Electronics
Circuits, Amplifiers and Gates

Second Edition

D V Bugg
Emeritus Professor, Queen Mary,
University of London

IoP
Institute of Physics Publishing
Bristol and Philadelphia

Copyright © 2005 IOP Publishing Ltd.

VISIT…

© IOP Publishing Ltd 2005

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission of the publisher. Multiple copying is permitted in accordance with the terms of licences issued by the Copyright Licensing Agency under the terms of its agreement with Universities UK (UUK).

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library.

ISBN 0 7503 1037 5

Library of Congress Cataloging-in-Publication Data are available

First edition 1991
Reprinted 1995, 1996, 1999

Commissioning Editor: Tom Spicer
Editorial Assistant: Leah Fielding
Production Editor: Simon Laurenson
Production Control: Sarah Plenty
Cover Design: Victoria Le Billon
Marketing: Louise Higham, Kerry Hopkins and Ben Thomas

Published by Institute of Physics Publishing, wholly owned by The Institute of Physics, London

Institute of Physics Publishing, Dirac House, Temple Back, Bristol BS1 6BE, UK

US Office: Institute of Physics Publishing, The Public Ledger Building, Suite 929, 150 South Independence Mall West, Philadelphia, PA 19106, USA

Typeset by Domex e-Data Pvt. Ltd.
Printed in the UK by MPG Books Ltd, Bodmin, Cornwall

Copyright © 2005 IOP Publishing Ltd.

Contents

	Preface		**ix**
1	**Voltage, Current and Resistance**		**1**
	1.1	Basic Notions	1
	1.2	Waveforms	3
	1.3	Ohm's Law	6
	1.4	Diodes	6
	1.5	Kirchhoff's Laws	8
	1.6	Node Voltages	14
	1.7	EARTHS	15
	1.8	Superposition	16
	1.9	Summary	19
2.	**Thevenin and Norton**		**24**
	2.1	Thevenin's Theorem	24
	2.2	How to Measure V_{EQ} and R_{EQ}	28
	2.3	Current Sources	30
	2.4	Norton's Theorem	31
	2.5	General Remarks on Thevenin's Theorem and Norton's	34
	2.6	Matching	35
	2.7	Amplifiers	36
	2.8	Systems	38
	2.9	Summary	38
3.	**Capacitance**		**44**
	3.1	Charge and Capacitance	44
	3.2	Energy Stored in a Capacitor	46
	3.3	The Effect of a Dielectric	46
	3.4	Capacitors in Parallel	47
	3.5	Capacitors in Series	48
	3.6	The CR Transient	48
	3.7	AC Coupling and Baseline Shift	53
	3.8	Stray Capacitance	55
	3.9	Integration and Differentiation	55
	3.10	Thevenin's Theorem Again	56
	3.11	Summary	57
4.	**Alternating Current (AC); Bandwidth**		**61**
	4.1	Introduction	61
	4.2	Power in a Resistor: RMS Quantities	61

Copyright © 2005 IOP Publishing Ltd.

	4.3	Phase Relations	62
	4.4	Response of a Capacitor to AC	64
	4.5	Simple Filter Circuits	65
	4.6	Power Factor	70
	4.7	Amplifiers	71
	4.8	Bandwidth	73
	4.9	Noise and Bandwidth	73
	4.10	Summary	75
5.	**Inductance**		**78**
	5.1	Faraday's Law	78
	5.2	Self-inductance	79
	5.3	LR Transient	80
	5.4	Energy Stored in an Inductor	82
	5.5	Stray Inductance	83
	5.6	Response of an Inductor to Alternating Current	84
	5.7	Phasors	84
	5.8	Summary	85
6.	**Complex Numbers: Impedance**		**89**
	6.1	Complex Numbers	89
	6.2	AC Voltages and Currents	93
	6.3	Inductance	95
	6.4	Summary on Impedance	96
	6.5	Impedances in Series	96
	6.6	Impedances in Parallel	98
	6.7	Power	101
	6.8	Bridges	102
7.	**Operational Amplifiers and Negative Feedback**		**109**
	7.1	Introduction	109
	7.2	Series Voltage Feedback	111
	7.3*	Approximations in Voltage Feedback	115
	7.4	Shunt Feedback	115
	7.5	The Analogue Adder	118
	7.6	The Differential Amplifier	119
	7.7	Gain-Bandwidth Product	120
	7.8	Offset Voltage and Bias Current	123
	7.9	Complex Feedback Loops	125
	7.10	Impedance Transformation	127
	7.11	Input and Output Impedances with Feedback	128
	7.12	Stabilised Current Supplies	132
	7.13*	Input Impedance with Shunt Feedback	133
	7.14	Oscillation	135

Copyright © 2005 IOP Publishing Ltd.

8. Integration and Differentiation **143**

8.1 Integration 143
8.2 The Miller Effect 146
8.3 Compensation 148
8.4 Differentiation 149
8.5 The Charge Sensitive Amplifier 149

9. The Diode and the Bipolar Transistor **155**

9.1 Conductors 155
9.2 Semiconductors and Doping 157
9.3 The pn Junction Diode 160
9.4 The Diode as a Switch 167
9.5 The npn Bipolar Transistor 169
9.6 Simple Transistor Circuits 171
9.7 Voltage Amplification 173
9.8 Biasing 175

10. The Field Effect Transistor (FET) **180**

10.1 Gate Action 180
10.2 Simple FET Amplifiers 183
10.3 MOSFETs 186
10.4 Fabrication of Transistors and Integrated Circuits 188
10.5 CMOS 190

11. Equivalent Circuits for Diodes and Transistors **192**

11.1 Introduction: the Diode 192
11.2 An Equivalent Circuit for the Bipolar Transistor 194
11.3 The Hybrid-π Equivalent Circuit 199
11.4 The FET 200
11.5 The Common Emitter Amplifier 201
11.6 Performance of the Common Emitter Amplifier 202
11.7 Emitter Follower 206
11.8 FETs 210

12. Gates **219**

12.1 Introduction 219
12.2 Logic Combinations of A and B 220
12.3 Boolean Algebra 223
12.4 De Morgan's Theorems 224
12.5 The Full Adder 225
12.6 The Karnaugh Map 226
12.7 Don't Care or Can't Happen Conditions 230
12.8 Products of Karnaugh Maps 231
12.9 Products of Sums 232
12.10 Use of NOR and NAND Gates 233

Copyright © 2005 IOP Publishing Ltd.

12.11 Decoders and Encoders 235
12.12 Multiplexing 237

13. Sequential Logic 241

13.1 The RS Flip-Flop 241
13.2 Clocks 242
13.3 The JK Flip-Flop 244
13.4 A Scale-of-4 Counter 246
13.5 State Diagrams 249
13.6 Trapping Sequences: Pattern Recognition 252
13.7 The Monostable 254
13.8 The Pulse Generator 256

14. Resonance and Ringing 261

14.1 Introduction 261
14.2 Resonance in a Series LCR Circuit 261
14.3 Transient in a CL Circuit 265
14.4 Transient in the Series LCR Circuit 266
14.5 Parallel LCR 269
14.6* Poles and Zeros 273

15. Fourier's Theorem 277

15.1 Introduction 277
15.2 A Square Wave applied to a CR Filter 278
15.3 How to Find Fourier Coefficients 279
15.4 The Heterodyne Principle 281
15.5 Broadcasting 282
15.6 Frequency Modulation (FM) 284
15.7 Frequency Multiplexing 285
15.8 Time Division Multiplexing 286
15.9 Fourier Series using Complex Exponentials 288
15.10 Fourier Transforms 289
15.11 Response to an impluse 290
15.12 Fourier analysis of a Damped Oscillator 291
15.13 The Perfect Filter 292

16. Transformers and 3-Phase Supplies 298

16.1 Introduction 298
16.2 Energy Stored in a Transformer 300
16.3 Circuit Equations and Equivalent Circuits 301
16.4 Three Phase Systems 304
16.5 Balanced Loads 306

Appendix A: Thevenin's Theorem 313

Appendix B: Exponentials 315

Copyright © 2005 IOP Publishing Ltd.

Preface

Industry and research depend heavily on electrical instruments and techniques. Trouble-shooting invariably involves using multimeters and oscilloscopes. For many students, experience with these instruments begins in first year courses at Universities and Colleges and thereafter continues to expand for many years. The objective of this book is to accompany first-year courses and go *just* beyond that for completeness. The aim is a rigorous but introductory treatment. I have frequently taught courses for engineers and have included the material they need - for example 3-phase power supplies. In the present perilous situation in which many departments find themselves, closer cooperation between engineering and physics faculties may be the way forward. I therefore hope the book may be helpful to both physicists and engineers.

Over 90% of the material included here has been used in practical classes. The necessary kit consists of a 2-beam oscilloscope, a breadboard, a power supply with two independent outputs up to 12 V, and a generator which provides sine waves and pulse trains up to $\sim 10^6$ Hz. Computer packages for Fourier analysis and Fourier transforms are a great help too. Student projects beyond the basic coursework are to be encouraged. They have a knack of uncovering knotty problems.

The new Edition is a slimmed down version of the first. Integrated circuits can now be bought so cheaply that they replace attempts at do-it-yourself design of electronics. This has made the more advanced parts of the earlier Edition redundant. However, I cling to the idea that all students need practical experience of how transistors amplify currents and voltages in a few rudimentary examples. The volume has been kept as slim as possible so that students can afford it, and can also carry it around!

Exercises graded in difficulty are provided with most chapters. Students are encouraged to check solutions in the lab. Many exercises come from examination papers set in colleges of London University, and I am grateful to the University for permission to reproduce them. Answers are mine and have been multiply checked, though I would be grateful to hear about any surviving errors.

Within the text, tricky questions are posed every now and then in italics. No answers are given, deliberately. I find that such unanswered teasers provoke as much thought and discussion as detailed exercises and I consider them an integral part of the learning process.

For me, it has been a lifetime of fun tackling physics problems and I hope this volume will be some return to the community, encouraging new students to follow a similar path. I owe a great debt to my wife for enduring the tribulations of doing physics.

David Bugg
February 2005

Copyright © 2005 IOP Publishing Ltd.

1

Voltage, Current and Resistance

1.1 Basic Notions

A thunderstorm is a dramatic electrical spectacle. Who could doubt the reality of voltages and currents when confronted by such a display? The storm separates electrical charges between top and bottom of the cloud. The energy required for this separation comes from warm moist air which rises and condenses. In a well developed storm, a voltage of ≈200 million volts builds up between the cloud and the earth. Eventually the air breaks down when the electric field at the base of the cloud approaches 10^6 V m^{-1}. In the ensuing lightning stroke the current is 10^4–10^5 amps.

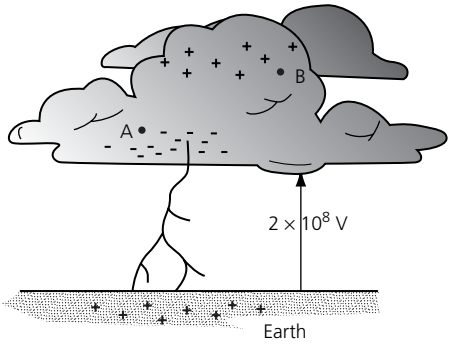


Fig. 1.1. Thundercloud and lightning.

The physics of a thunderstorm is complicated. For present purposes this example serves to focus attention on basic notions of charge, current, voltage and energy. What do the numbers mean? How are voltage and current defined?

First let's get clear the relation between **voltage** and **energy**. If a charge Q is moved from point A in the cloud at voltage V_A to point B at voltage V_B, (figure 1.2),

Copyright © 2005 IOP Publishing Ltd.

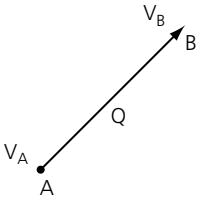


Fig. 1.2. Work done $= Q(V_B - V_A)$.

the work done on the charge is $Q \times (V_B - V_A)$. In SI units, it requires one joule(J) to move one coulomb(C) of charge through one volt (V):

$$\boxed{\text{Energy (J)} = \text{Charge (C)} \times \text{Potential difference (V)}.} \tag{1.1}$$

When the thundercloud discharges, this energy is liberated as heat, thunder and electromagnetic radiation. If the cloud carries a charge of 10 C, the energy liberated in a complete discharge is $10 \times 2 \times 10^8$ J = 2×10^9 J or 2 GJ (Gigajoules).

In an electrical circuit, a battery likewise provides a potential difference and a source of energy. A fully charged 12 V car battery can deliver a charge of typically 3×10^5 C from chemical reactions and an energy of 3.6×10^6 J or 3.6 MJ (megajoules).

A third familiar example will illustrate these ideas. Figure 1.3 shows schematically the layout of an oscilloscope. Electrons are emitted from a heated cathode and are then accelerated through a large voltage V, typically 1–5 kV (kilovolts). They acquire kinetic energy in falling through the potential difference V and reach a velocity v given by

$$eV = \tfrac{1}{2}mv^2 \tag{1.2}$$

where e and m are the charge and mass of the electron. We must distinguish between the voltage difference V through which the beam is accelerated and the energy eV acquired by each electron. If $V = 5$ kV, this energy is $1.6 \times 10^{-19} \times 5 \times 10^3$ J = 8×10^{-16} J.

Next, what is the relation between current and charge? The moving electrons carry a current, whose magnitude is defined by the charge Q passing through the control grid in unit time: $1\,Ampere = 1\,Coulomb\ per\ second$. Since the current may vary with time, it needs to be expressed in terms of differential quantities:

$$\mathrm{d}Q(C) = I(A) \times \mathrm{d}t(s)$$

or

$$\boxed{I = \mathrm{d}Q/\mathrm{d}t.} \tag{1.3}$$

For example, a beam of 10^{13} electrons/s carries a current of $-1.6 \times 10^{-19} \times 10^{13}$ A $= -1.6 \times 10^{-6}$ A or -1.6 μA (microamps). The minus sign arises because

Copyright © 2005 IOP Publishing Ltd.

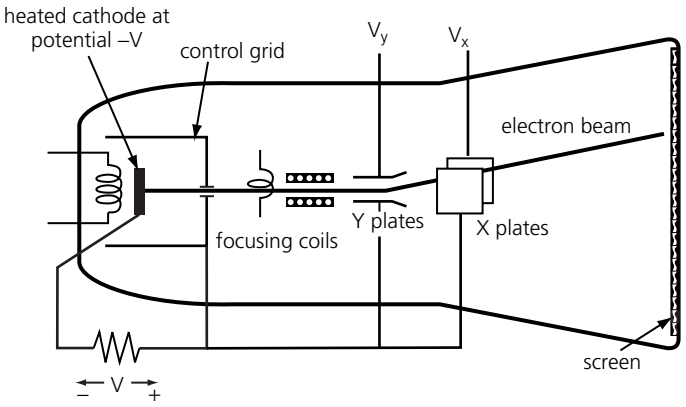


Fig. 1.3. Schematic of an oscilloscope.

electrons are negatively charged; conventional current flows in the opposite direction to the electrons. The current may be measured using the force on it in a magnetic field. You can demonstrate this force by holding a bar magnet up to an oscilloscope or a TV set and watching the beam deflect. The *Ampere*, the absolute unit of current, is actually defined in terms of the force between two current-carrying coils. Then the *Coulomb* is derived from the *Ampere* using equation (1.3).

Another important electrical quantity is **power** P. It is defined as the rate of change of energy, E:

$$P = \mathrm{d}E/\mathrm{d}t. \tag{1.4}$$

It may be related to voltage and current using equation (1.1). A charge $\mathrm{d}Q$ falling through potential V gains energy $\mathrm{d}E = V\mathrm{d}Q$, so a current I flowing through potential difference V generates power

$$\boxed{P = V\frac{\mathrm{d}Q}{\mathrm{d}t} = VI.} \tag{1.5}$$

This is the familiar result for power dissipated in a resistor. Power is measured in watts(W): one watt = 1 joule/second. In each stroke of a lightning discharge, power is dissipated for $\approx$100 μs at a rate of about 5×10^{12} W. By comparison, a large power station generates 2000 MW = 2×10^9 W.

1.2 Waveforms

This chapter is concerned mostly with constant voltages and currents. Such a situation is referred to as DC, meaning **direct current**. However, the principles

Copyright © 2005 IOP Publishing Ltd.

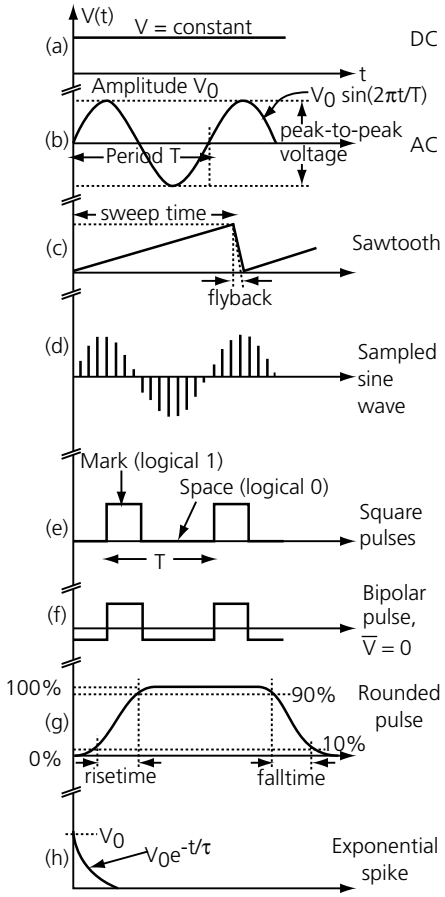


Fig. 1.4. Common waveforms.

carry over to situations where currents and voltages are varying. In the simplest case, figure 1.4(b), they vary with time t as $\sin \omega t$ or $\cos \omega t$. Current like this is called AC or **alternating current**. Although AC and DC ought to apply strictly to current, they are often used loosely for voltages too.

Waveforms are readily displayed on the screen of an oscilloscope. Technically, this is done as follows. The beam of electrons of figure 1.3 is focussed by electrostatic or magnetic lenses. Two pairs of plates provide electric fields at right angles to the beam; they deflect it horizontally(X) and vertically(Y). (In a TV set, the plates are replaced by magnetic coils.) The ramp or sawtooth voltage of figure 1.4(c) is applied to the horizontal plates, while the waveform $V(t)$ is applied to the vertical plates. The result is a graph of $V(t)$ against t, figure 1.5. When the display leaves the screen, the sawtooth voltage V_0 is restored rapidly to zero. During this 'flyback', the source is blanked off, to avoid confusing the picture.

Copyright © 2005 IOP Publishing Ltd.

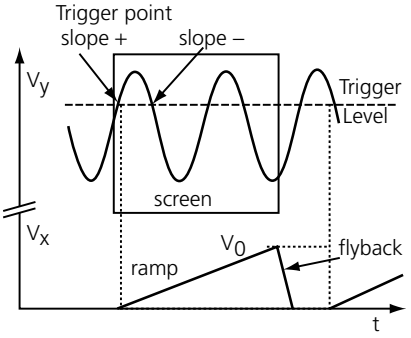


Fig. 1.5. Triggering of an oscilloscope.

The virtue of the oscilloscope is its speed. Common oscilloscopes display waveforms on a timescale of ≈1 s/cm to 1 μs/cm; high quality oscilloscopes using sampling techniques can display features down to 10^{-10} s or less.

Several other common waveforms are shown in figure 1.4. Square pulses are used in digital circuitry. It is arbitrary whether a low voltage stands for binary 0 and a high voltage for binary 1 or vice versa. Bipolar pulses are square pulses superimposed on a DC level which makes the mean voltage zero averaged over time. (This avoids charging capacitors, see Chapter 3). Spikes are used to trigger digital circuits.

In circuit diagrams, generators which produce sine waves, square waves, or ramps are drawn as in figure 1.6. The appropriate waveform is shown inside the circle representing the generator. Where it matters, the polarity of the signal is indicated explicitly, as in (d). Strictly speaking, a battery should be represented

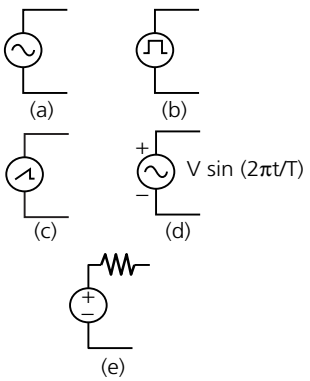


Fig. 1.6. Symbols for generators of sine waves, square pulses and ramps; (e) represents a battery explicitly.

Copyright © 2005 IOP Publishing Ltd.

as in figure 1.6(e) by a DC generator with a resistance in series, to represent the internal resistance of the battery. This formality can frequently be skipped, but you should remember that batteries always have some internal resistance. Indeed, all generators have some internal resistance called the output impedance.

1.3 Ohm's Law

Experimentally, the current in many materials, such as metals, depends linearly on applied voltage. This is Ohm's law. The constant of proportionality is defined so that

$$\boxed{V = IR.} \tag{1.6}$$

The **resistance** R is measured in ohms (Ω) if V is in volts and I in amps. For wires of a particular material, it is found that the resistance is proportional to the length L of the wire and inversely proportional to its area A:

$$R = \rho L/A \tag{1.7}$$

where ρ is the resistivity of the material. Can you justify equation (1.7) in terms of resistors in series and parallel? The best conductors are silver and copper with $\rho = 1.5$ and $1.7 \times 10^{-8}\,\Omega$ m respectively, not very dependent on temperature.

Do you know what your resistance is from one hand to the other? Try measuring it with a multimeter or AVOmeter, (Amps, Volts, Ohms - meter). First touch the leads lightly with one finger of each hand, then make contact by pressing each finger against a coin touching the meter lead. Your resistance depends on the area of contact. If you moisten your fingers the resistance decreases, but then the resistance of the rest of the body dominates.

The meter works off a 6 or 9 V battery. Estimate roughly what current flows through your body. Mild shocks are felt from a current of 1 mA, currents of 10 mA are dangerous and 100 mA may be fatal. To what voltages do these correspond? When you work with high voltages, make a habit of using only one hand and keep the other in your pocket. That eliminates the risk of a current from hand to hand, affecting the heart, although there is still the danger of a current from your hand to your feet, if they make good electrical contact with the ground. Make sure you do not grasp anything at high voltage, so that if you do get a shock you jump clear.

1.4 Diodes

For a resistor, V increases linearly with I, figure 1.7(a). However, not all materials obey Ohm's law. In a lightning discharge, current varies non-linearly with current. A second example is provided by the *pn* diode, where the relation between current

Copyright © 2005 IOP Publishing Ltd.

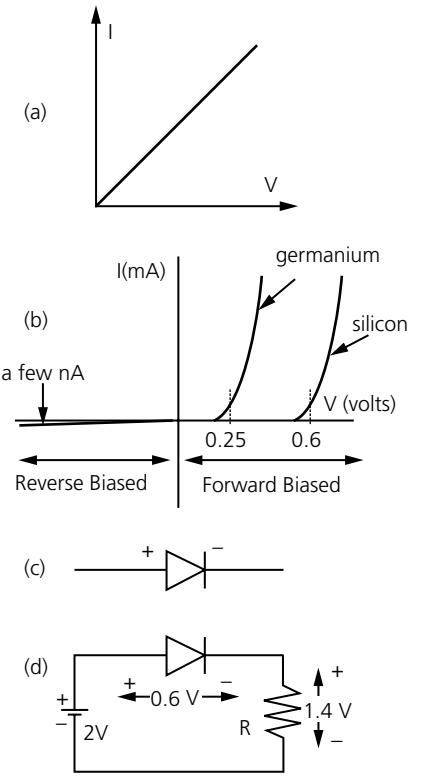


Fig. 1.7. Relation between I and V for (a) a resistor, (b) diodes, (c) the symbol for the diode, (d) a circuit including a diode.

and voltage is sketched in figure 1.7(b). Current flows much more easily in one direction than the other. When it flows easily the diode is said to be **forward biased** and when the voltage is the other way it is said to be **reverse biased**. The symbol for the diode is shown in figure 1.7(c); the arrow denotes the direction in which the diode conducts.

We shall find later that the current in a diode rises exponentially with voltage for forward bias. Typical applications use currents of 1–100 mA. For silicon diodes, the current rises steeply through this range as V increases from about 0.55 to 0.65 V. When the diode conducts to a significant extent, the voltage across it is usually close to 0.6 V; this voltage may be regarded as the value required to switch the diode on. For lower or negative voltages, little current flows. In the circuit of figure 1.7(d), the diode is forward biased and the current is governed by the voltage $(2.0 - 0.6)$ V = 1.4 V across the resistor R.

For reverse bias, the diode current is very small, a few nA for silicon, and one can usually think of the diode as non-conducting. Diodes are rated according to the reverse voltage they stand before breaking down. It is possible to control the

Copyright © 2005 IOP Publishing Ltd.

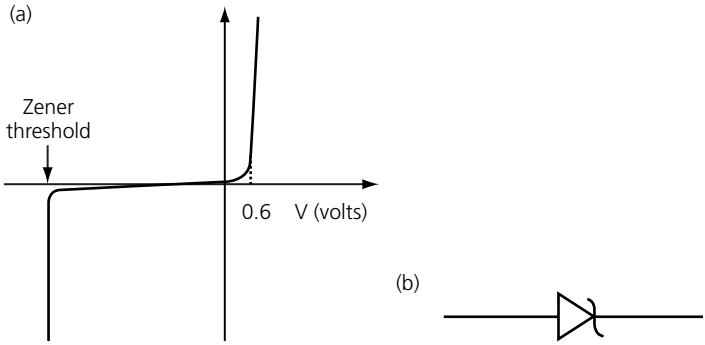


Fig. 1.8. (a) Zener diode characteristics, (b) the symbol for the Zener diode.

chemical doping of the material so that this breakdown occurs at a precisely controlled voltage which can provide a reference value in electrical circuits. This kind of diode is called a **Zener diode**. Its characteristic curve is shown in figure 1.8(a) and its circuit symbol in figure 1.8(b); again the arrow indicates the direction of current flow when it is forward biased, although it is mostly used reverse biased, to make use of the threshold. Typical values of this threshold for reverse bias are 2.7 to 30 V.

> *How can you distinguish between a deviation from Ohm's law and a non-linearity in the meters you use for measuring voltage and current? As a clue, consider how you can double a voltage and how you can double a current.*

1.5 Kirchhoff's Laws

The remainder of the chapter develops the methods needed to calculate currents and voltages round a circuit, e.g. figure 1.9(a). There are several methods and short-cuts. Problems like this crop up endlessly in practical work, so you need to be able to solve these problems fluently.

Kirchhoff's laws are the foundation. They work for any components: resistors, capacitors, inductors, diodes and transistors, whether or not Ohm's law is obeyed. They reduce any circuit, however complicated, to a set of simultaneous equations. The rest is algebra. Sometimes it is convenient to express the equations in terms of currents, sometimes in terms of voltages. We shall explore both.

Kirchhoff's laws express two fundamental laws of physics: conservation of charge and conservation of energy. The first is **Kirchhoff's current law**. At any junction or **node**, charge is conserved and so is current, figure 1.9(b); remember $I = \mathrm{d}Q/\mathrm{d}t$. This conservation law allows us to reduce the number of variables by using the currents I_4 and I_5 which flow round closed loops of figure 1.9(a). These are called **mesh currents**. They are related to I_{1-3} by

Copyright © 2005 IOP Publishing Ltd.

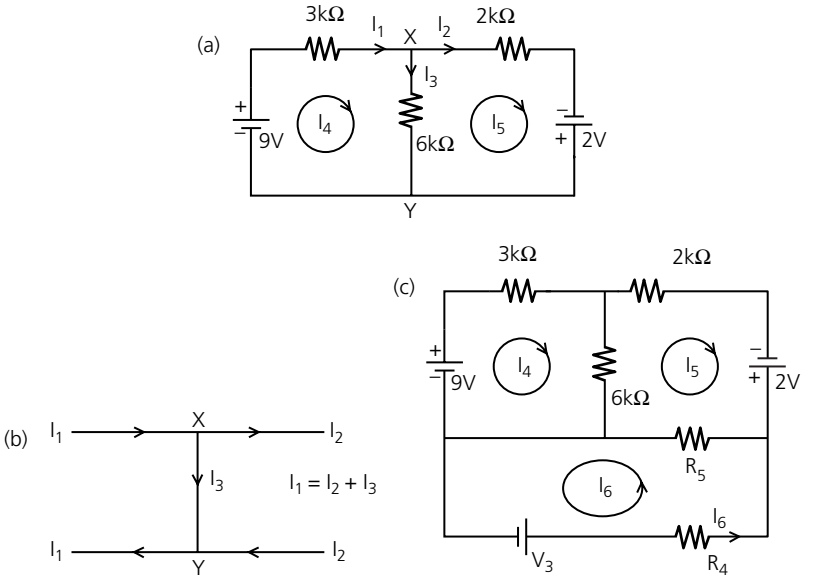


Fig. 1.9. (a) Illustrating Kirchoff's laws, (b) current conservation, (c) 3 loops.

$I_1 = I_4$; $I_2 = I_5$ and $I_3 = I_4 - I_5$. Using mesh currents automatically disposes of Kirchhoff's current law.

Next we apply **Kirchhoff's voltage law**. Round any closed loop of the circuit, energy conservation requires that the net change of potential is zero. The changes in potential over individual components of a loop sum to zero. Care is required over signs: which voltage changes are positive and which negative. If in figure 1.9(a) we start at the point Y and move clockwise round each loop, the voltages across each component give

$$9 - 3I_4 - 6(I_4 - I_5) = 0$$

$$6(I_4 - I_5) - 2I_5 + 2 = 0$$

where I is in mA (milliamps). There are two equations for the two unknowns I_4 and I_5 with the solution $I_4 = 7/3$ mA and $I_5 = 2$ mA. After solving the equations, it is a good idea to evaluate the potential drops across individual resistors as a check, particularly on signs.

If more loops containing resistors and batteries are added as in figure 1.9(c), each new loop adds one further current but one further equation. For figure 1.9(c), the equations are

$$9 = 3I_4 + 6(I_4 - I_5) \qquad \text{as before}$$

$$2 = R_5(I_5 + I_6) + 6(I_5 - I_4) + 2I_5$$

$$V_3 = R_4 I_6 + R_5(I_5 + I_6).$$

Copyright © 2005 IOP Publishing Ltd.

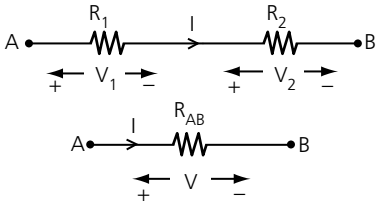


Fig. 1.10. Resistors in series

Kirchhoff's laws may be used to find equations for networks of any degree of complexity, though solving the equations may be tedious. There are many computer packages to help do this algebra.

Series resistors

The familiar formulae for resistances in series and parallel will be derived, as elementary examples of Kirchhoff's laws. In figure 1.10, the voltages and resistances are **in series**. Kirchhoff's voltage law gives

$$V = V_1 + V_2 = IR_1 + IR_2 = I(R_1 + R_2).$$

The relation between V and I is just the same as for an equivalent resistor R_{AB} where

$$\boxed{R_{AB} = R_1 + R_2 \qquad \text{(series)}.} \tag{1.8}$$

This equivalence also holds for power dissipation:

$$VI = I^2R_1 + I^2R_2 = V^2/(R_1 + R_2)$$

and this is equal to the power dissipated in R_{AB}, namely V^2/R_{AB}.

The potential divider

A common arrangement of resistors is the potential divider, figure 1.11(a). Suppose nothing is applied externally between X and Y. Then V_X is fixed at a value intermediate between V and V_Y:

$$V = V_1 + V_2 = I(R_1 + R_2)$$

$$\boxed{V_{XY} = IR_2 = VR_2/(R_1 + R_2).} \tag{1.9}$$

Copyright © 2005 IOP Publishing Ltd.

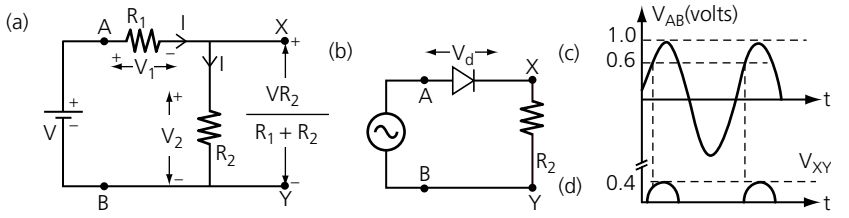


Fig. 1.11. (a) A potential divider, (b) a diode instead of R_1, (c) and (d) waveforms for (b).

You can use this result to verify Ohm's law. Suppose the battery is replaced by a sine wave generator. The equation works for AC voltages as well as DC: the applied voltage $V(t)$ varies with time, but at any instant Ohm's law still applies. If you display V_{XY} and V_{AB} on two traces of an oscilloscope, you can verify that V_{XY} is a scaled down replica of V_{AB}. Most oscilloscopes have a knob allowing you to vary continuously the vertical scale of each trace, so with a little dexterity you can superpose one waveform on top of the other. This demonstrates that $V = IR$.

If however you replace R_1 with a diode, as in figure 1.11(b), and set the amplitude of the sine wave generator to $\sim$ 1 V, you get a result more like that in figure 1.11(c); V_{XY} is quite different in shape to V_{AB}. When the diode conducts, the voltage V_d across it is roughly 0.6 V and $V_{XY} = V_{AB} - V_d$; when $V_{XY} < 0.6$ V, the diode does not conduct, so $V_{XY} \simeq 0$. Ohm's law fails here because the diode behaves non-linearly.

You may have noticed that the oscilloscope has a resistance of typically $10^6\ \Omega = 1\ \text{M}\Omega$. This is called its **input resistance**, R_{in}, or **input impedance**. If this is similar to R_2, you must allow for it. It is a good idea to measure the resistances of the oscilloscope and multimeter you use in the lab, so that you can allow for them when necessary. This is easily done with the circuit of figure 1.12 by measuring the current $I = V/R_{in}$. When you use a multimeter to measure a resistance, the circuit is identical to figure 1.12, with an internal battery driving current through the meter and resistance. If the resistor is in a circuit, you need to disconnect one end to make sure that (a) other components in the circuit do not appear in parallel, (b) voltage sources in the circuit do not drive current through the meter and (c) the battery in the meter does not affect the way the circuit operates.

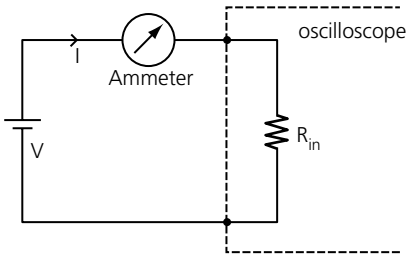


Fig. 1.12. Measuring R_{in}

Copyright © 2005 IOP Publishing Ltd.

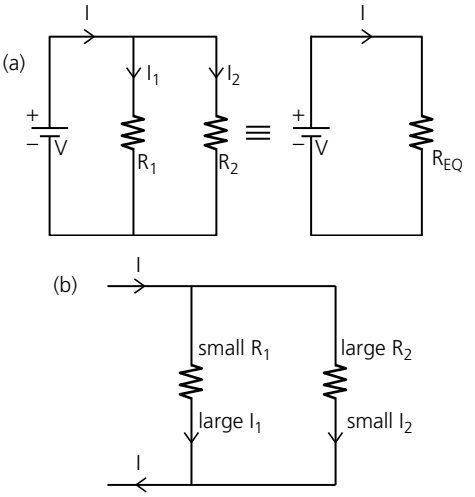


Fig. 1.13. Resistors in parallel.

Parallel resistors

As another exercise, the formulae for resistors in parallel, figure 1.13, will be derived. The same potential V appears across R_1 and R_2, so

$$V = I_1 R_1 = I_2 R_2. \tag{1.10}$$

By Kirchhoff's current law, the total current is

$$I = I_1 + I_2 = (V/R_1) + (V/R_2) \equiv V/R_{EQ}. \tag{1.11}$$

Hence

$$\boxed{\frac{1}{R_{EQ}} = \frac{1}{R_1} + \frac{1}{R_2} = \frac{R_2 + R_1}{R_1 R_2}} \tag{1.12}$$

or

$$R_{EQ} = \frac{R_1 R_2}{R_1 + R_2} \qquad \text{(parallel)}. \tag{1.12a}$$

Both forms need to be memorised.

It is useful to make approximations if the resistors have widely different values. Suppose R_2 is much the larger. It carries little current and $R_{EQ} \simeq R_1$. More exactly,

$$R_{EQ} = \frac{R_1}{1 + R_1/R_2} \simeq R_1(1 - \frac{R_1}{R_2}) \tag{1.12b}$$

Copyright © 2005 IOP Publishing Ltd.

where the binomial theorem has been used to expand $(1+R_1/R_2)^{-1}$. For example, if $R_1 = 1\ \text{k}\Omega$ and $R_2 = 33\ \text{k}\Omega$, $R_{EQ} \simeq 1 \times (1 - 0.03)\ \text{k}\Omega = 0.97\ \text{k}\Omega$. Get used to doing this rough arithmetic.

Parallel resistors act as a **current divider** and it will be useful to memorise expressions for I_1 and I_2. From (1.10) and (1.11):

$$I = I_1(1 + R_1/R_2),$$

$$\boxed{I_1 = \frac{I R_2}{R_1 + R_2} \qquad \text{and} \qquad I_2 = \frac{I R_1}{R_1 + R_2}.} \tag{1.13}$$

The larger current flows in the smaller resistor, figure 1.13(b). Again the total power dissipated in R_1 and R_2 is the same as in R_{EQ}:

$$P = V I_1 + V I_2 = \frac{V^2}{R_1} + \frac{V^2}{R_2} = \frac{V^2}{R_{EQ}}.$$

Worked example

It is often easy to reduce a network by combining resistors in series and parallel. Make sure you solve for the quantity you want to know. Suppose, for example, you require the total power dissipated in the circuit of figure 1.14(a). You need

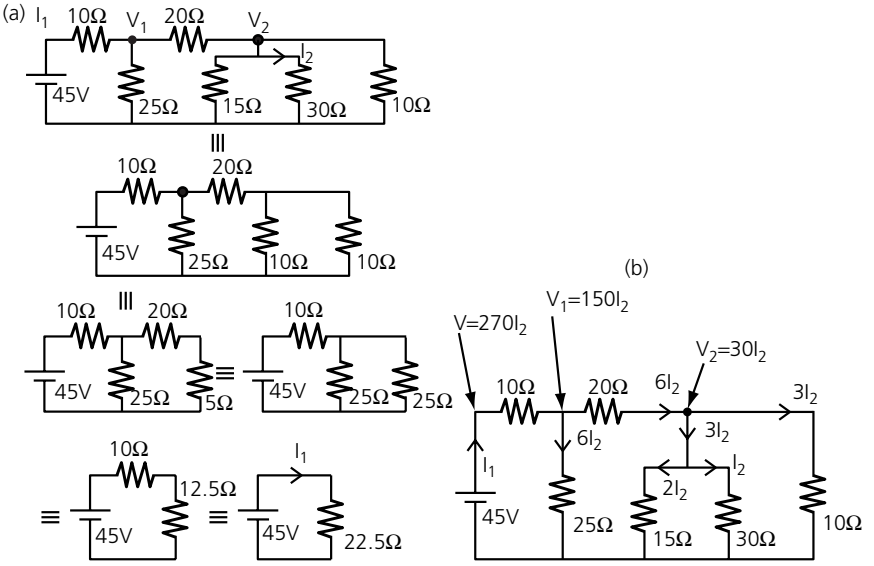


Fig. 1.14. (a) Reduction of a network by series and parallel equivalents, (b) solution in terms of I_2.

Copyright © 2005 IOP Publishing Ltd.

to find the current I_1 delivered by the battery. This is easily done by the steps shown in the figure. From the final diagram, $I_1 = 2$ A and the power supplied is $P = 45 \times 2$ W = 90 W.

Suppose, however, you want to know instead the current I_2 in the 30 Ω resistor and the power dissipated there. Instead of solving for I_1 and then back-tracking to I_2, it is convenient to evaluate all voltages and currents in terms of I_2 from the outset. This is illustrated in figure 1.14(b). Firstly $V_2 = 30I_2$. Currents in the 15 Ω and 10 Ω resistors are $2I_2$ and $3I_2$, so the total current through the 20 Ω resistor is $6I_2$ and

$$V_1 = V_2 + 120I_2 = 150I_2.$$

This procedure continues in the same way back to the battery, which is known to be 45 V. The current through the 25 Ω resistor is $6I_2$ and finally $I_1 = 12I_2$; so the voltage at the battery is given by $V_1 + 120I_2 = 270I_2$. Hence $I_2 = 1/6$ A. The power in the 30 Ω resistor is $V_2 I_2^2 = 30I_2^2 = 5/6$ W.

Examples at the end of the chapter develop the tricks you frequently need. Note, however, that it is not always possible to reduce networks to series and parallel combinations. Figure 1.9(c) is a counter-example. There, you must use Kirchhoff's laws directly.

Although using Kirchhoff's laws is a safe method, it is often long-winded. In the example we have just done, blind application of Kirchhoff's laws leads to four mesh currents and four simultaneous equations. The remainder of this chapter and the next will explore shortcuts giving the answer more directly.

1.6 Node Voltages

In the worked example, it came in useful to choose the voltages V_1 and V_2 at nodes as unknowns. This often reduces the number of equations. Figure 1.9(a) serves as an example. There are two nodes X and Y. Suppose voltages are measured relative to the point Y by taking $V_Y = 0$. This leaves only one unknown, the voltage V_X at X. The key point is to evaluate currents at this node in terms of V_X and use the fact that currents sum to zero at the node. The current through the 3 kΩ resistor

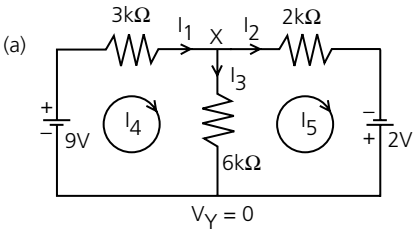


Fig. 1.9. (a) again.

Copyright © 2005 IOP Publishing Ltd.

is $(9 - V_X)/3$ mA, that through the 6 kΩ resistor is $(V_X/6)$ mA and the current through the 2 kΩ resistor is $(V_X + 2)/2$ mA. Then

$$\frac{9 - V_X}{3} = \frac{V_X}{6} + \frac{V_X + 2}{2}$$

or $V_X = 2$ V. From this, I_5 and I_4 follow immediately; the results of course agree with the previous solution. In this method, only a single equation has to be solved instead of the two simultaneous equations of the previous section. This is a useful simplification.

1.7 EARTHS

Only voltage differences appear in Kirchhoff's laws. The absolute voltages of points in the circuit are irrelevant. In fact, one of the subtleties of electromagnetism is that absolute voltages can *never* be measured.

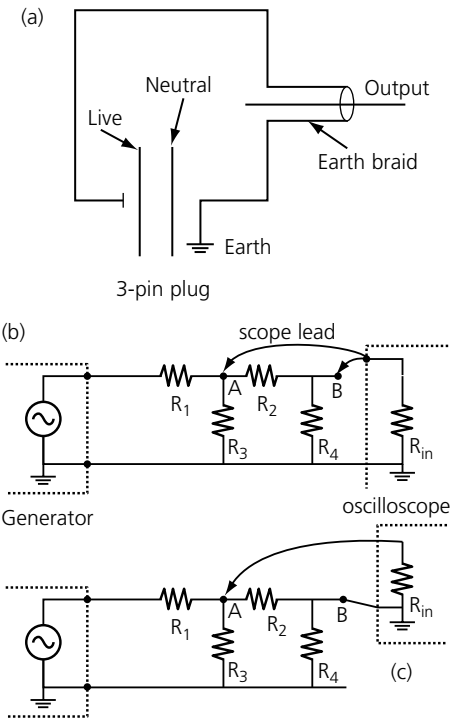


Fig. 1.15. (a) Earthing convention, (b) correct and (c) incorrect ways to connect earth leads when measuring a circuit.

Copyright © 2005 IOP Publishing Ltd.

It is very convenient in analysing a circuit to define $V = 0$ at some point such as Y of the previous section. In practical work, it is likewise convenient to measure voltages from that of the Earth. Figure 1.15(a) shows a widespread convention in electrical equipment. The earth terminal of the 3-pin plug is connected internally to the metal instrument case and any earth terminals on it; they are all at the same potential called **earth** or **ground**. This way, anything you touch should be incapable of giving you a shock. The earthing point is denoted by the symbol shown in figure 1.15(a). It will be explained later that the **neutral** differs from Earth by at most 1 or 2 V under all ordinary circumstances. The **live** wire carries high voltage from the mains.

Output and input signals are developed between the central wire and earth. When you connect instruments to a circuit, e.g. a sine wave generator and an oscilloscope, you need to remember this convention and connect their earths to a common point, as in figure 1.15(b). If instead you mistakenly connect one of the earth terminals to a point which is not meant to be earthed, as in figure 1.15(c), you short out part of the circuit via the mains leads and upset its operation. If you want to find the signal across R_2 using an oscilloscope, you will need to make separate measurements at points A and B and take the difference. Oscilloscopes with two traces make this easy.

For this particular measurement, a multimeter may be more convenient. Its case is not earthed, but is left **floating**. It can measure the voltage *difference* between any two points in a circuit.

1.8 Superposition

Let us return to the circuit of figure 1.9(a), reproduced in figure 1.16 and work out what happens if each battery in turn is reduced to zero, but otherwise the circuit is left connected. The currents for these two cases are shown in (b) and (c).

If we add the results for (b) and (c), the total currents are the same as in (a), i.e. the currents produced by the individual batteries simply add. This is the **Superposition Principle**. It is a consequence of the linear relation between voltage and current, and holds in general when network elements obey this linear relation. It implies that the currents produced by one battery split at nodes in a way which depends only on the resistances in the network, independently of other batteries. This can be handy (a) in thinking about the way circuits behave, (b) in doing arithmetic on individual circuits. For example, in dealing with figure 1.9(c), we wrote down three simultaneous equations, but were then faced with the algebraic problem of solving them. It may be more convenient to find the currents generated by each of the three batteries in turn. Further examples are given in the worked example which follows and in exercises at the end of the chapter.

Copyright © 2005 IOP Publishing Ltd.

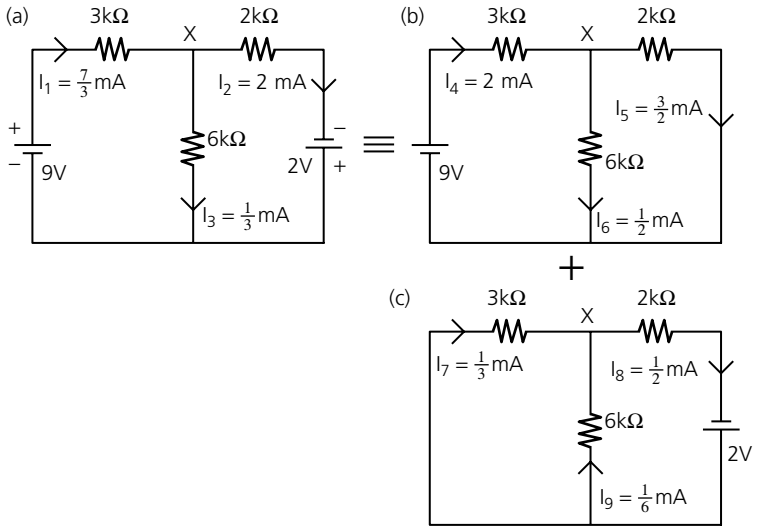


Fig. 1.16. The superposition principle.

A nice demonstration is to replace one battery of figure 1.16 by a sine wave generator. Although the applied voltage varies with time, the superposition principle applies at any instant. The voltage across any resistor may be displayed on the oscilloscope, and is a superposition of a DC voltage from the battery and an AC component from the generator, figure 1.17. Running the generator down to zero leaves the DC component unchanged; taking out the battery leaves the AC component unchanged.

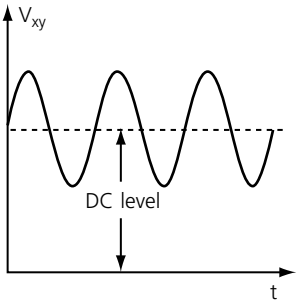


Fig. 1.17. Superposition of DC and AC voltages.

Another worked example

Figure 1.18(a) shows a rather complicated circuit. As an illustration, we shall find all the currents and the voltages V_A and V_B at nodes by our three different methods.

Copyright © 2005 IOP Publishing Ltd.

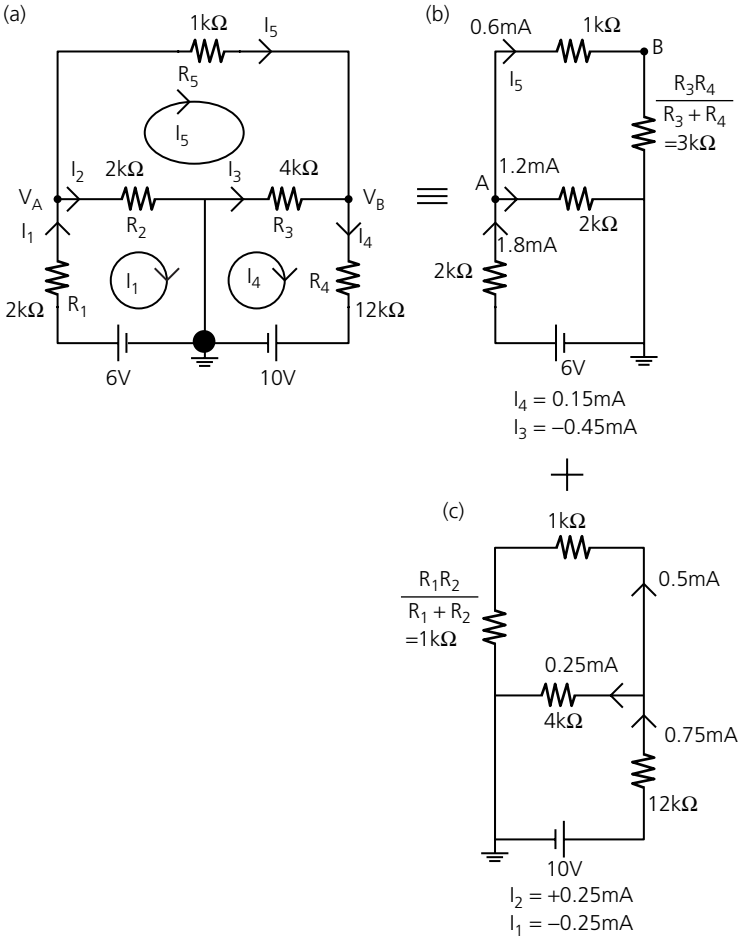


Fig. 1.18. Worked example.

Using superposition, the currents may be obtained from those due to individual batteries. In (b), the 10 V battery is shorted out and R_3 and R_4 appear in parallel. The top three resistors of (b) provide 2 kΩ in parallel with 4 kΩ between A and earth, i.e. $(4/3)$ kΩ, so the contribution to $I_1 = 1.8$ mA. It is easy to see how this divides at A; then I_5 splits between 0.15 mA through R_4 and a contribution of -0.45 mA to I_3. In (c), the 6 V battery is shorted, so R_1 and R_2 appear in parallel. The arithmetic of the resulting currents is shown in the figure. The signs of the contributions to I_1 and I_2 are easy to follow from (a) and the sense in which the 10 V battery drives currents. Adding currents from (b) and (c), $I_1 = 1.55$ mA, $I_2 = 1.45$ mA, $I_3 = -0.7$ mA, $I_4 = -0.6$ mA and $I_5 = 0.1$ mA. From these currents, it is simple to find $V_A = 6 - 3.1 = 2.9$ V and $V_B = 10 + I_4 R_4 = 2.8$ V.

Copyright © 2005 IOP Publishing Ltd.

Suppose instead the problem is to be solved using mesh currents. The ones to choose would be I_1, I_4 and I_5, as shown in (a). Applying Kirchhoff's current law, $I_1 = I_2 + I_5$ and $I_4 = I_3 + I_5$. The values given in the previous paragraph satisfy these relations. Then applying Kirchhoff's voltage law to each loop in turn:

$$6 = 2I_1 + 2(I_1 - I_5) = 4I_1 - 2I_5$$
$$10 = -12I_4 + 4(I_5 - I_4) = 4I_5 - 16I_4$$
$$0 = 1I_5 + 4(I_5 - I_4) + 2(I_5 - I_1) = 7I_5 - 4I_4 - 2I_1$$

with currents in mA. Solving these three simultaneous equations is tedious. It is however straightforward to substitute the values derived above and demonstrate that the equations are correctly satisfied. Using superposition is really a graphical way of eliminating variables from the simultaneous equations.

The third alternative is to use node voltages V_A and V_B. Then current conservation at these nodes gives

$$\frac{6 - V_A}{2} = \frac{V_A - V_B}{1} + \frac{V_A}{2}$$
$$\frac{10 - V_B}{12} = \frac{V_B}{4} + \frac{V_B - V_A}{1}.$$

The solution of these two equations is easy; a check is that the equations are satisfied by the values of V_A and V_B obtained above.

A warning

Superposition is a valuable shortcut, but it only works exactly for circuits containing linear components like resistors, where $V \propto I$. The next chapter develops other powerful shortcuts which again depend on linearity. However, many electronic devices such as diodes and transistors do not obey Ohm's law, but follow a non-linear relation between current and voltage like figure 1.7(b). This problem will be postponed to Chapter 11.

1.9 Summary

Kirchhoff's voltage and current laws apply to any circuit, whether components are linear or not. Series resistors make a potential divider, parallel resistors a current divider. Networks may be solved using mesh currents or node voltages; either may be convenient under different circumstances, and it is a matter of practice to spot which gives the easier solution. When the circuit components are linear, superposition offers a shortcut; it fails however when the voltage-current relation for any component moves over a significantly non-linear region.

Copyright © 2005 IOP Publishing Ltd.

1.10 Exercises

It is desirable to do enough examples to become fluent at calculating currents and voltages. Those marked with one asterisk are lengthy or tricky. Those marked with two asterisks are revealing and not necessarily difficult.

1. Write notes on the meaning of voltage, energy, current and resistance. In Fig. 1.11(a), what is the voltage at X if point A is earthed?
2. *Series resistance and power.* The lights on a Christmas tree are 20×1 W bulbs in series connected to the 240 V mains. One bulb fuses, but still conducts. What current flows and what power is now dissipated in each bulb? (Ans: 88 mA, 1.11 W.)
3. *Parallel resistances.* Two resistors R_1 and R_2 in parallel have an equivalent resistance of 30 Ω. Current divides between them in the ratio 3:1. Determine R_1 and R_2. (Ans: 40 Ω and 120 Ω.)
4. *Approximations.* A 33 kΩ resistor is in parallel with 330 kΩ. Find their equivalent resistance (a) exactly, (b) using the binomial theorem and equation (1.12b). (Ans: (a) 30 kΩ, (b) 29.7 kΩ.)

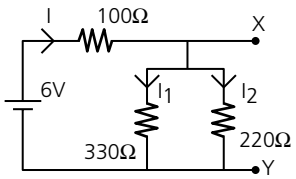


Fig. 1.19.

5. *Potential and current dividers.* Find V_{XY} in figure 1.19 and currents I_1 and I_2. (Ans: $V_{XY} = 3.41$ V, $I_1 = 10.34$ mA, $I_2 = 15.5$ mA.)

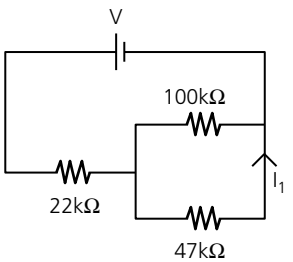


Fig. 1.20.

Copyright © 2005 IOP Publishing Ltd.

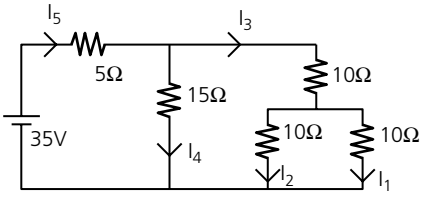


Fig. 1.21.

6. *Parallel resistances and currents*. In the circuit in figure 1.20, $I_1 = 50\ \mu$A. What is V? What current flows through the 100 kΩ resistor? (Ans: 3.97 V; 23.5 μA.)

7. *Parallel resistances and power*. Determine the currents I_1–I_5 in figure 1.21. Find the total power loss in the resistors and verify that it equals the power delivered by the battery. (Ans: $I_1 = I_2 = 0.7$ A; $I_3 = I_4 = 1.4$ A; $I_5 = 2.8$ A; $P = 98$ W.)

8. *Instrument impedance*. Figures 1.22(a) and (b) show two ways of measuring R. The voltmeter V has a resistance of 20 kΩ and the ammeter A a resistance of 100 Ω. If $V_{IN} = 12$ V and $R = 4.7$ kΩ, what current and voltage are measured in each case? (Ans: (a) 12 V, 2.5 mA, (b) 11.69 V, 3.07 mA.)

9. *Mesh currents*. Use mesh currents in the circuit of figure 1.20 to find I_1 and the current supplied by the battery if $V = 10$ V. Check by combining resistors in series and parallel. (Ans: $I_1 = 126$ mA, 185 mA.)

10**. *Mesh currents and Node voltages*. Find the node voltages V_1 and V_2 in figure 1.23. (Remember that this method depends on currents adding at nodes). Hence find the mesh currents I_1, I_2 and I_3 and check that your solution satisfies Kirchhoff's voltage law round each loop. (Ans: $V_1 = 35/6$ V, $V_2 = 5$ V; $I_1 = 15/144$ A; $I_2 = 45/144$ A; $I_3 = 115/144$ A.)

11**. *Superposition*. Find I_1 and I_2 in figure 1.24 due to battery V_1 acting alone and also due to V_2 acting alone. Show that the sum of these currents satisfies Kirchhoff's voltage law for the loops. (Ans: V_1 alone gives $I_1 = V_1(R_2 + R_3)/X$, $I_2 = -V_1R_2/X$, where $X = R_1R_2 + R_1R_3 + R_2R_3$; V_2 alone gives $I_1 = -V_2R_2/X$, $I_2 = V_2(R_1 + R_2)/X$.)

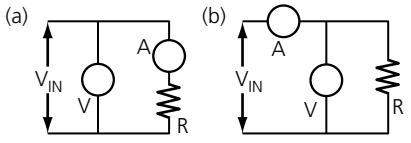


Fig. 1.22.

Copyright © 2005 IOP Publishing Ltd.

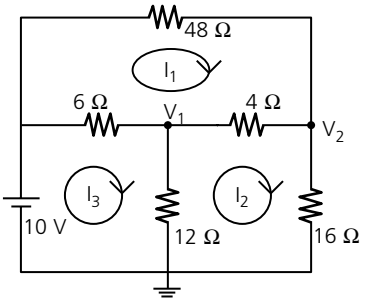


Fig. 1.23.

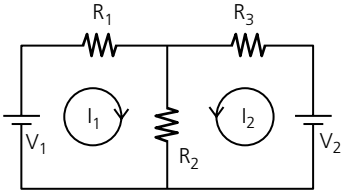


Fig. 1.24.

12*. *Node voltage and superposition*. Find the voltage V_X at node X. Then find the currents I_1–I_4 in figure 1.25. Check these currents using superposition of the currents due to the 1 V and 2 V batteries; (note that in doing this, the battery not being used must be replaced by a short circuit). (Ans: $V_X = 1.75$ V; $I_1 = 30$ mA, $I_2 = 17.5$ mA, $I_3 = -7.5$ mA, $I_4 = 25$ mA.)

13. *Attenuator, Characteristic Impedance*. The network shown in figure 1.26 attenuates an input voltage V_{in}, producing an output voltage $V_{out} = V_{in}/A$ across a load Z. The value of n can be chosen so that

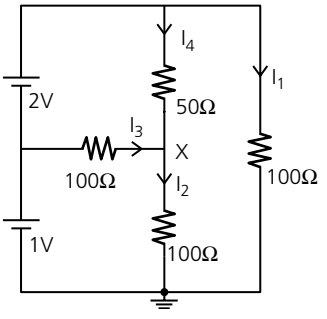


Fig. 1.25.

Copyright © 2005 IOP Publishing Ltd.

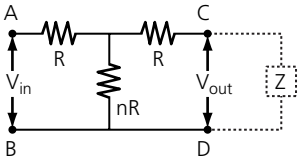


Fig. 1.26.

the impedance measured across the input terminals AB is also Z when the load is connected. In this case, the network is said to have **characteristic impedance** Z; then m elements like figure 1.26 can be joined in sequence to make a ladder network with (a) input impedance Z which is independent of m and (b) output signal V_{in}/A^m across the terminating load Z. Each step of the ladder from AB to CD acts as an **attenuator**, reducing the signal by a constant factor A. Show that the required value of n is $0.5(x^2 - 1)$ where $x = Z/R$ and that A is then $(x + 1)/(x - 1)$. Find n and R if $Z = 100\ \Omega$ and A is (a) 2, (b) 10. Comment on the stability required for R in the two cases. (Ans: (a) $n = 4$, $x = 3$, (b) $n = 20/81$, $x = 11/9$; A sensitive to n and R in the latter case.)

Copyright © 2005 IOP Publishing Ltd.

2

Thevenin and Norton

2.1 Thevenin's Theorem

We now come to a very important method which will be extended later to circuits containing capacitors, inductors and transistors. It is important in two quite different respects: firstly, it is often a shortcut, secondly, it leads to conceptual simplifications in thinking about complicated circuitry.

Often what matters is only the current or voltage in one component of a circuit. It is a waste of time solving for all currents or voltages, and mistakes can easily arise in solving many equations. Thevenin's theorem bypasses this drudgery. It will be introduced by means of a simple example, shown in figure 2.1(a).

Suppose we want to know the current I_2 between the terminals AB through a load resistor R_L. Thevenin's theorem states that I_2 is given by the equivalent circuit of figure 2.1(b). The circuit enclosed by the dashed lines behaves simply as a source of voltage V_{EQ} in series with an internal resistance R_{EQ}. A straightforward way of deriving V_{EQ} and R_{EQ} will emerge below. Then the voltage V_{AB} across the load resistor is

$$\boxed{V_{AB} = V_{EQ} - R_{EQ} I_2 = I_2 R_L.} \tag{2.1}$$

This theorem works no matter how complicated the network of batteries and resistors inside the dashed lines, with the one proviso that all the elements obey a linear relation between voltage and current. That is the case for circuits containing just batteries and resistors. If all the components in the circuit are linear, it is very reasonable that V_{AB} should be linearly related to I_2. What is less obvious is that the values of V_{EQ} and R_{EQ} depend only on the values of components within the dashed lines *and are independent of* R_L *and hence* I_2. For practical purposes, the dashed lines could represent a black box, with two output terminals; all that matters are V_{EQ} and R_{EQ}. For a complicated circuit, this is a great conceptual and practical simplification. Let us first check it for the simple circuit of figure 2.1(a).

Copyright © 2005 IOP Publishing Ltd.

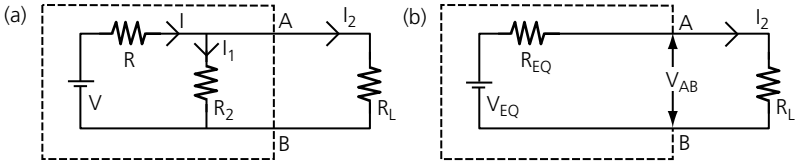


Fig. 2.1. (a) Load across a potential divider, (b) the Thevenin equivalent circuit.

To do this, the equations describing figure 2.1(a) need to be arranged so as to display current I_2 and voltage V_{AB}:

$$I_2 = I - I_1 = \frac{V - V_{AB}}{R_1} - \frac{V_{AB}}{R_2} = \frac{V}{R_1} - V_{AB}\left(\frac{1}{R_1} + \frac{1}{R_2}\right).$$

Rearranging,

$$V_{AB} = \frac{V R_2}{R_1 + R_2} - I_2 \frac{R_1 R_2}{R_1 + R_2}.$$

This indeed looks like equation (2.1) if

$$V_{EQ} = \frac{V R_2}{R_1 + R_2} \tag{2.2}$$

$$R_{EQ} = \frac{R_1 R_2}{R_1 + R_2}. \tag{2.3}$$

Notice that V_{EQ} and R_{EQ} do not depend on R_L, but only on quantities within the dashed lines.

This example verifies Thevenin's theorem in one particular case. The general case, where there is a large number of loops in the network, goes the same way and is given in Appendix A.

Finding V_{EQ} and R_{EQ}

Assuming Thevenin's theorem, there is a quicker way to arrive at V_{EQ} and R_{EQ}. Remember that these two quantities are independent of R_L. It is convenient to consider three special cases. Any two of them give V_{EQ} and R_{EQ} and the third acts as a cross-check against mistakes.

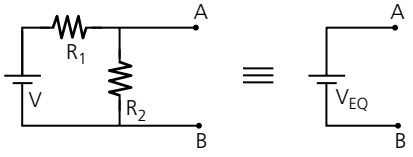


Fig. 2.2. Finding V_{EQ}.

Copyright © 2005 IOP Publishing Ltd.

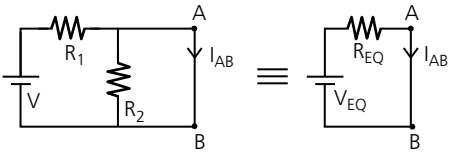


Fig. 2.3. Shorting AB.

i) If $R_L \to \infty$ (i.e. an open circuit across AB), no current flows through AB, figure 2.2, so there is no voltage drop in R_{EQ}; then

$$V_{EQ} = V_{AB} = V R_2/(R_1 + R_2)$$

in agreement with (2.2). The value of V_{EQ} is equal to V_{AB} when AB is open circuit.

ii) If terminals AB are short-circuited (figure 2.3), no current flows through R_2, so $I_{AB} = V/R_1$. But in this case the short circuit current I_{AB} is equal to V_{EQ}/R_{EQ}, hence

$$R_{EQ} = V_{EQ}/I_{AB} = \frac{V R_2}{R_1 + R_2} \times \frac{R_1}{V} = \frac{R_1 R_2}{R_1 + R_2},$$

in agreement with (2.3).

iii) Imagine that the voltage V is scaled down by a large factor f. All the currents in the circuit scale down by this factor; resistances remain unchanged in both figure 2.1(a) and its equivalent circuit. Ultimately, as $V \to 0$, the equivalent circuit contains just R_{EQ} to the left of the terminals AB, figure 2.4(b). Thus R_{EQ} is found by reducing all voltage sources to zero, i.e. by short-circuiting them. Figure 2.4(a) contains R_1 and R_2 in parallel to the left of AB:

$$R_{EQ} = \frac{R_1 R_2}{R_1 + R_2}$$

in agreement with (ii) and (2.3).

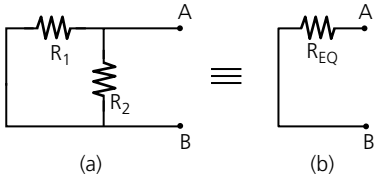


Fig. 2.4. Finding R_{EQ}.

Copyright © 2005 IOP Publishing Ltd.

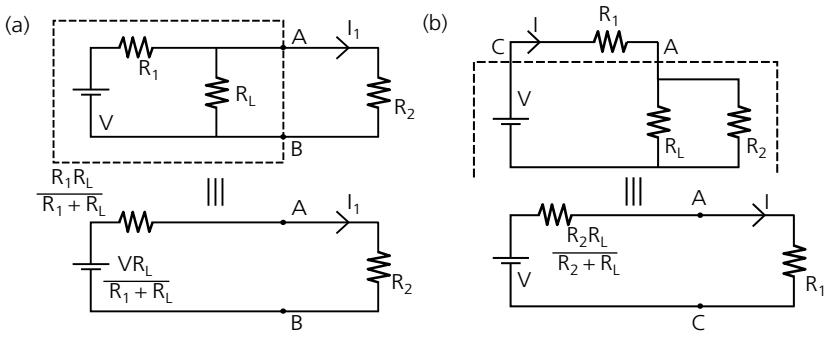


Fig. 2.5. Equivalent circuits across (a) R_2, (b) R_1.

Method (iii) is applicable only if voltage sources in the circuit are independent of current, as in the example discussed above. Later we shall encounter voltage sources (and current sources) whose magnitude is proportional to another voltage or current in the circuit. In this case, it is no longer permissible to scale voltages and currents down to zero and method (iii) no longer works, though (i) and (ii) do. We shall return to this refinement later.

A second warning is that the equivalent circuit constructed here refers specifically to the terminals AB across resistor R_L. The equivalent circuit across R_1 or R_2 will be different. Suppose, for example, figure 2.1 is redrawn as in figure 2.5(a) to display the equivalent circuit across R_2; it is an exercise for the student to show that the new equivalent circuit has $R_{EQ} = R_1 R_L/(R_1 + R_L)$ and $V_{EQ} = V R_L/(R_1 + R_L)$. These have different values to those for the previous equivalent circuit across R_L.

Examples

A familiar example of a Thevenin equivalent circuit is that for a battery: a voltage source with an internal resistance. The concept is however quite general. A pulse generator is a complicated box of electronics; but it behaves just like a voltage source $V_{EQ}(t)$ with an output resistance R_{EQ}. The voltage source may have a square wave dependence on time t or some other t dependence (e.g. a

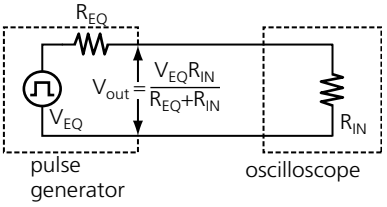


Fig. 2.6. Electronic systems represented by their Thevenin equivalents.

Copyright © 2005 IOP Publishing Ltd.

sawtooth). However, providing the circuit is linear, R_{EQ} does not vary with V_{EQ} (or equivalently with current). If it does, the circuit is non-linear. Even if R_{EQ} varies slightly with current, the approximate notion of a Thevenin equivalent circuit may be a useful guide.

Suppose the pulse generator is connected to an oscilloscope, figure 2.6. The oscilloscope may again be viewed in terms of its Thevenin equivalent, which is just an input resistance. In this way a very complicated set of circuits has been reduced to two simple black boxes. A manufacturer will normally quote the output or input resistance of his product in its specification. For most applications, he goes to considerable trouble to make the performance linear, so that Thevenin's theorem applies.

2.2 How to Measure V_{EQ} and R_{EQ}

Suppose we are presented with a gadget which measures the electrical voltage produced by a nerve cell. We know nothing about the internal workings of the gadget, only that it can be represented by a Thevenin equivalent circuit. How do we find V_{EQ} and R_{EQ}? If the gadget is connected to an oscilloscope, as in figure 2.6, the voltage measured by the oscilloscope is $V_{EQ}R_{IN}/(R_{IN} + R_{EQ})$. If $R_{IN} \gg R_{EQ}$, the oscilloscope measures V_{EQ} directly. Oscilloscopes usually have a large input resistance, typically 1 MΩ. Special probes are available to raise the input resistance to 10 or 20 MΩ if need be.

The value of R_{EQ} may be determined by substituting various loads R_L for the oscilloscope and measuring the output voltage V_{out}. The current I_L through the load is V_{out}/R_L, and should follow the straight line relation of figure 2.7(a). If not, the components inside the black box are behaving non-linearly, leading to a breakdown of Thevenin's theorem. The straight line on the figure is called the **load line** of the device, since it describes how it responds to a load. Its slope determines R_{EQ}.

Often R_{EQ} may be found by adjusting R_L until $V_{out} = V_{EQ}/2$, as in figure 2.7(b); in this case, $R_L = R_{EQ}$. However, be careful not to draw too large a current and damage the source; a car battery has a resistance which is a fraction of an ohm and the battery would be damaged by applying such a small resistance across the terminals.

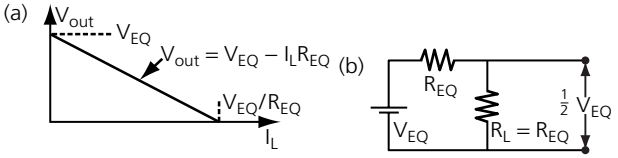


Fig. 2.7. (a) straight-line relation between V_{out} and I_L, (b) $V_{out} = V_{EQ}/2$ when $R_L = R_{EQ}$.

Copyright © 2005 IOP Publishing Ltd.

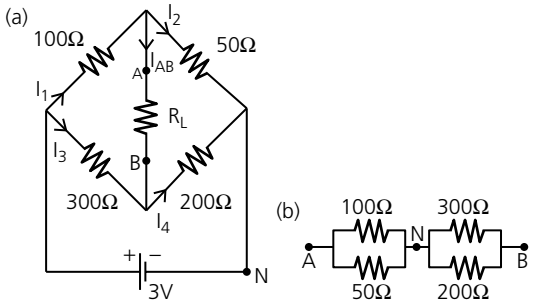


Fig. 2.8. (a) Wheatstone bridge, (b) R_{EQ}.

Second worked example on Thevenin's Theorem
Consider the Wheatstone bridge of figure 2.8. Suppose we want to find the current through the galvanometer, of resistance R_L, between A and B. It is necessary to construct the Thevenin equivalent circuit across terminals AB.

i) To find V_{EQ}, let $R_L \to \infty$. Taking the negative terminal N of the battery as the zero of voltage,

$$V_A = \frac{3 \times 50}{100 + 50} = 1 \text{ V}$$

$$V_B = \frac{3 \times 200}{200 + 300} = 1.2 \text{ V}.$$

$$V_{EQ} = V_A - V_B = -0.2 \text{ V}.$$

ii) To find R_{EQ}, suppose the battery is replaced by a short-circuit. Then the resistors between A and B may be redrawn as in figure 2.8(b), so

$$R_{EQ} = \frac{100 \times 50}{100 + 50} + \frac{300 \times 200}{300 + 200} = 33\tfrac{1}{3} + 120 = 153\tfrac{1}{3}\ \Omega.$$

iii) In this particular example, finding I_{AB} when $R_L = 0$ is a little lengthy, but still a worthwhile check. It is equal to $(I_1 - I_2)$ and also $(I_4 - I_3)$, so it is necessary to solve for these currents:

$$V_A = 50I_2$$

$$I_1 = (3 - V_A)/100 = 0.03 - 0.5I_2 = I_2 + I_{AB},$$

hence $$I_{AB} = 0.03 - 1.5I_2.$$

Copyright © 2005 IOP Publishing Ltd.

Also

$$I_4 = V_A/200 = I_2/4$$

$$I_3 = (3 - V_A)/300 = 0.01 - I_2/6.$$

$$I_3 = I_4 - I_{AB} = 1.75I_2 - 0.03.$$

Equating these two expressions for I_3,

$$0.04 = (I_2/12)(2 + 21) = 23I_2/12.$$

Finally

$$I_{AB} = (0.69 - 0.72)/23 = -0.03/23 \text{ A}.$$

This is consistent with V_{EQ}/R_{EQ}.

2.3 Current Sources

So far we have considered voltage sources such as batteries and waveform generators. The ideal battery produces a constant output voltage, regardless of current, as in figure 2.9(a). There do exist sources, however, which behave much more nearly like figure 2.9(b); the output current is constant, regardless of voltage. Such a source is called a **constant current source**. It is represented in circuit diagrams by the symbol of figure 2.9(c), for both AC and DC situations. An example is a photodiode described in Chapter 9. When light falls on this diode, electrons are liberated and the current through the diode is proportional to the light intensity.

Non-ideal voltage sources produce an output voltage which drops with current as in figure 2.7(a) and figure 2.9(d); for small output resistance R_{EQ}, the slope is small. A non-ideal current source also produces a diagram like figure 2.9(d), but with large slope (large output resistance). There is no clear distinction between non-ideal voltage and non-ideal current sources, and we shall now demonstrate that the circuits of figure 2.10 are interchangeable: they produce the same straight line relation between V_{AB} and I_L. The circuit within the dashed lines of figure 2.10(b) is Norton's equivalent circuit and is an alternative to Thevenin's circuit, figure 2.10(a).

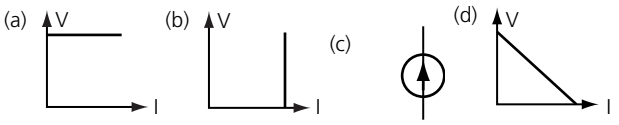


Fig. 2.9. (a) Ideal voltage source, (b) ideal current source, (c) the symbol for a current source, (d) non-ideal source.

Copyright © 2005 IOP Publishing Ltd.

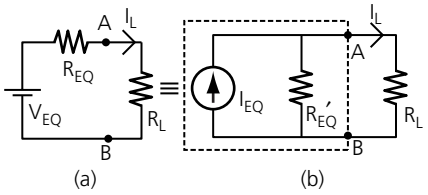


Fig. 2.10. Norton's equivalent circuit.

2.4 Norton's Theorem

Norton's theorem states that any network with two output terminals is equivalent to a current source I_{EQ} with R'_{EQ} in parallel, figure 2.10(b). As with Thevenin's theorem, it depends on components in the network obeying a linear relation between current and voltage. It follows very simply from Thevenin's theorem. Suppose in figure 2.10 current I_L flows through a load resistor R_L attached to the terminals AB. Then in the Norton circuit, current $(I_{EQ} - I_L)$ flows through R'_{EQ} and

$$V_{AB} = R'_{EQ}(I_{EQ} - I_L) = R'_{EQ}I_{EQ} - R'_{EQ}I_L.$$

For Thevenin's circuit, $V_{AB} = V_{EQ} - R_{EQ}I_L$. These two results agree if

$$\boxed{R'_{EQ} = R_{EQ}} \tag{2.4}$$

$$\boxed{V_{EQ} = R_{EQ}I_{EQ}.} \tag{2.5}$$

These two relations may be checked very simply from the fact that when AB is open circuit $V_{EQ} \equiv R'_{EQ}I_{EQ}$ and on short circuit $I_L = I_{EQ} \equiv V_{EQ}/R_{EQ}$.

The value of R_{EQ} in the Thevenin equivalent circuit was obtained from the internal resistance with the batteries short-circuited. This is clear from the left-hand side of figure 2.10. If there are current sources present, these are replaced by open circuits when finding R_{EQ}; this follows from an inspection of the right-hand side of figure 2.10.

Example using Norton's Theorem

i) In figure 2.11(a), if terminals AB are shorted, the current through them is 10/5 mA from the 10 V battery and 20/10 mA from the 20 V battery. So $I_{EQ} = 4$ mA.

ii) With the batteries replaced by short circuits, R_{EQ} is the resistance across AB, namely 5 kΩ in parallel with 10 kΩ. So $R_{EQ} = 10/3$ kΩ.

Copyright © 2005 IOP Publishing Ltd.

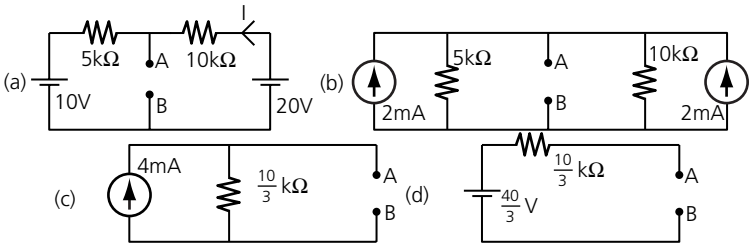


Fig. 2.11. (a) Example of Norton's equivalent circuit, (b)–(d) ways of simplifying it.

iii) As a check, consider the situation with AB open circuit. In this case, there is a net voltage of 10 V in (a), driving current I in the direction of the arrow through the 5 and 10 kΩ resistors; $I = 10/15$ mA and $V_{AB} = 10 + 5I = 10 + 10/3 = 40/3$ V. This agrees with $I_{EQ}R_{EQ}$ from (i) and (ii).

A circuit can often be simplified quickly and neatly by swopping backwards and forwards between Thevenin and Norton equivalent forms. This is illustrated in figures 2.11(b) – (d). It is a trick worth practising, since it often saves a great deal of algebra. The batteries and resistors of (a) are replaced by equivalent Norton circuits in (b); these are combined in parallel in (c) and then (d) converts back to the Thevenin equivalent form. An important warning is that you must not include the load resistor between terminals A and B in these manipulations: Thevenin's theorem and Norton's apply to the circuits *feeding* terminals AB.

Another worked example

Figure 2.12 reproduces a fairly complicated example from Chapter 1, figure 1.18(a). If all currents and voltages in the circuit are required, it is best to use one of the methods from Chapter 1. However, suppose only current I_2 is

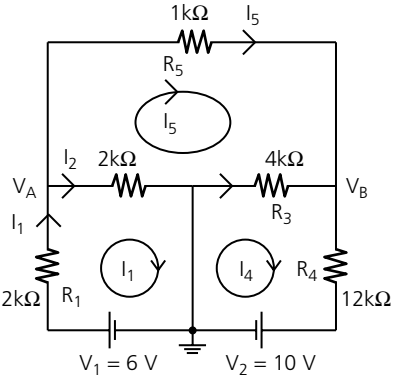


Fig. 2.12. A worked example.

Copyright © 2005 IOP Publishing Ltd.

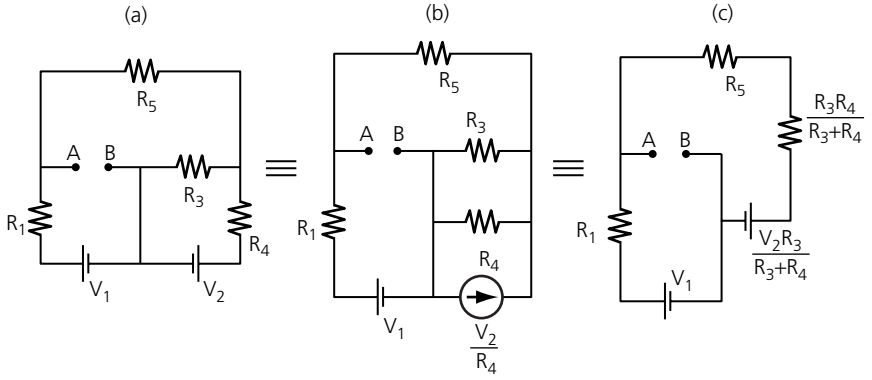


Fig. 2.13. Worked example.

needed. It can be obtained by applying Thevenin's theorem and Norton's. The steps are shown in figure 2.13. In (b), V_2 and R_4 are replaced by their Norton equivalent. Then R_3 and R_4 are combined in parallel and (c) returns to the Thevenin equivalent form.

With AB open circuit,

$$V_{EQ} = V_1 - \left(V_1 - \frac{V_2 R_3}{R_3 + R_4}\right) R_1 / \left(R_1 + R_5 + \frac{R_3 R_4}{R_3 + R_4}\right)$$

$$= 6 - (6 - 2.5)2/(2 + 1 + 3) = 29/6 \text{ V}.$$

With the batteries shorted out, R_{EQ} is given by the parallel combination of R_1 with

$$R_5 + R_3 R_4/(R_3 + R_4),$$

i.e. 2 kΩ in parallel with 4 kΩ; so $R_{EQ} = (4/3)$ kΩ.

As a check, the current through AB when shorted is I_{EQ}:

$$I_{EQ} = \frac{V_1}{R_1} + \frac{V_2 R_3}{R_3 + R_4} / \left(R_5 + \frac{R_3 R_4}{R_3 + R_4}\right) = 3 + 2.5/4 = 29/8 \text{ mA}.$$

This agrees with V_{EQ}/R_{EQ} as it should. The arithmetic and algebra is sufficiently tortuous that this is a valuable crosscheck.

Finally the current I_2 of figure 2.12 is

$$I_2 = V_{EQ}/(R_{EQ} + R_2) = \tfrac{29}{6} / \left(\tfrac{4}{3} + 2\right) = 1.45 \text{ mA}$$

in agreement with the value obtained in the previous chapter.

Copyright © 2005 IOP Publishing Ltd.

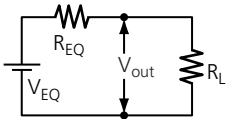


Fig. 2.14. For constant V_{out}, $R_{EQ} \ll R_L$.

Further examples are given in the exercises at the end of the chapter. If you can do question 6, you have mastered the vital points of Chapters 1 and 2 up to here.

2.5 General Remarks on Thevenin's Theorem and Norton's

1) Suppose a constant voltage is required across a load R_L, with as little variation as possible when R_L is changed. From figure 2.14, R_{EQ} needs to be small compared with R_L, so that most of V_{EQ} appears across R_L. Thus a constant voltage source should have a low output resistance.

2) Conversely, suppose a constant output current is required, independent of load; this is the case in supplying a magnet or a motor. From figure 2.15, this demands $R_{EQ} \gg R_L$ or high output resistance.

3) When a circuit is measured with an oscilloscope or voltmeter, it is desirable to disturb the circuit as little as possible, i.e. draw very little current. The detector must have a high input resistance or **input impedance**. Oscilloscopes have input resistances of 10^6–10^7 Ω. On the other hand, if an ammeter is inserted into a circuit in order to measure current, we want to disturb the current as little as possible. So an ammeter should have a low resistance.

4) Although Thevenin's circuit and Norton's are equivalent in the sense of giving the same output voltage and current, they are NOT equivalent as regards power consumption *within* the equivalent circuit. This is because power is nonlinear in V or I. You may easily verify that the power dissipated inside the Norton equivalent circuit of figure 2.10(b) is different from that in the Thevenin equivalent circuit (a).

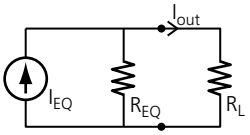


Fig. 2.15. For constant I_{out}, $R_{EQ} \gg R_L$.

Copyright © 2005 IOP Publishing Ltd.

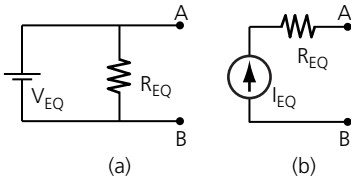


Fig. 2.16. Wrong choices for equivalent circuits.

5) Common student howlers are to draw equivalent circuits in the forms shown in figure 2.16. It is worth a moment's thought as to why these must be wrong. In the former case, $V_{AB} = V_{EQ}$ independent of load; this gives an absurd result if the terminals are shorted. In the second circuit, $I_{AB} = I_{EQ}$ independent of load; this is absurd if the terminals are open.

6) If you encounter a circuit like that in figure 2.16(a) where a resistor is applied directly across a battery, you can ignore the resistor in forming an equivalent circuit, since V_{AB} is equal to the voltage of the battery, regardless of the resistor value. The resistor dissipates power but has no effect on the equivalent circuit. In a circuit like figure 2.16(b), $I_{AB} = I_{EQ}$ regardless of the resistor value.

2.6 Matching

Suppose a given voltage source has fixed characteristics V_{EQ} and R_{EQ}. The question may arise how to deliver maximum power to a load R_L, figure 2.17(a). Suppose there is the freedom to vary R_L. The current flowing into the load is

$$I = V_{EQ}/(R_{EQ} + R_L)$$

and the power P dissipated in the load is

$$P = I^2 R_L = V_{EQ}^2 R_L/(R_{EQ} + R_L)^2.$$

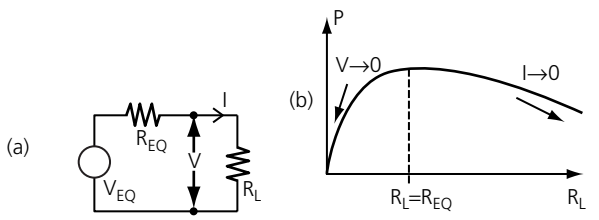


Fig. 2.17. (a) A load R_L connected to a voltage source, (b) P as a function of R_L.

Copyright © 2005 IOP Publishing Ltd.

This is zero when $R_L = 0$ or when $R_L \to \infty$. In between, there must be a maximum, figure 2.17(b). This requires $\mathrm{d}P/\mathrm{d}R_L = 0$. Now

$$\frac{\mathrm{d}P}{\mathrm{d}R_L} = \frac{V_{EQ}^2}{(R_{EQ}+R_L)^2} - \frac{2V_{EQ}^2 R_L}{(R_{EQ}+R_L)^3} = \frac{V_{EQ}^2(R_{EQ}-R_L)}{(R_{EQ}+R_L)^3}$$

so the power is a maximum when $R_L = R_{EQ}$. In this case, R_{EQ} is said to be **matched** to the load R_L.

There is however a snag with this arrangement. Half of the power is dissipated in the source itself. This is wasteful. If R_{EQ} can be varied, it is still desirable to minimise it, so that most of the power is dissipated in the load.

2.7 Amplifiers

This section takes a brief look forward at transistor amplifiers where Thevenin and Norton equivalent circuits play a vital role. An amplifier is an essential element of any electronic system. It amplifies voltage or current or both. In Chapter 11, it will be shown that a voltage amplifier may be represented by the circuit of figure 2.18(a).

Internally, the amplifier may be very complicated. Whatever the complexity, the input and output may be represented by Thevenin equivalent circuits, providing voltage and current are linearly related. In reality, the back-EMF V_{back} in the input

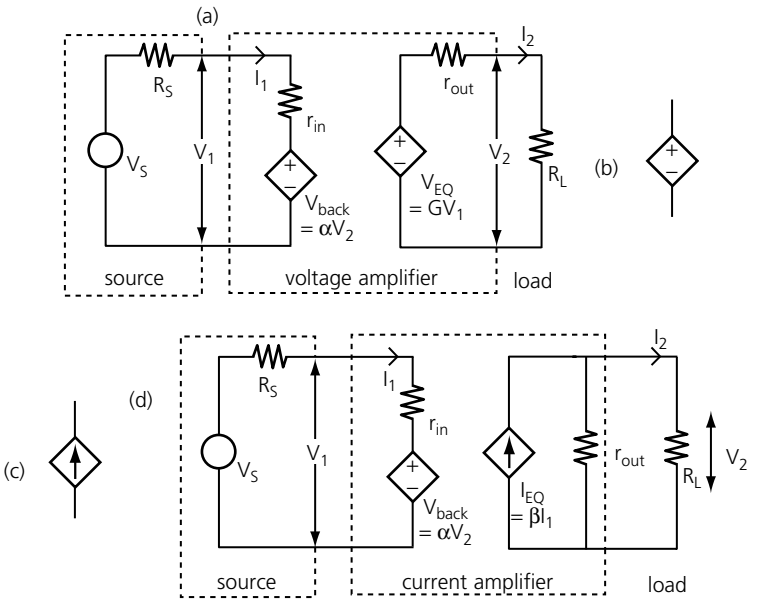


Fig. 2.18. (a) A voltage amplifier, (b) a dependent voltage source, (c) a dependent current source, (d) a current amplifier.

Copyright © 2005 IOP Publishing Ltd.

side of the amplifier is often negligible; then the input behaves simply as an input resistance r_{in}.

The output circuit consists of a voltage generator $V_{EQ} = GV_{in} = GV_1$ amplifying the input voltage, together with an impedance r_{out}. The voltage generator is called a **dependent voltage source** because it depends on another voltage. The symbols for a dependent voltage source and a dependent current source are shown in figures 2.18(b) and (c). The circuit of figure 2.18(a) obeys the equations

$$V_2 = GV_1 - r_{out}I_2 \tag{2.17a}$$

$$V_1 = r_{in}I_1 + \alpha V_2. \tag{2.17b}$$

For efficient coupling of source to load, r_{in} should be large compared with R_S; and r_{out} should be small compared with R_L. Then $V_1 \simeq V_S$ and $V_2 \simeq V_{EQ}$.

A current amplifier instead amplifies current, figure 2.18(d). This is obtained from figure 2.18(a) simply by replacing the output Thevenin equivalent circuit by the Norton form. Technically there is no difference between these two forms. Which is more convenient depends on whether G is approximately constant as operating conditions vary or β. For figure 2.18(d),

$$I_2 = \beta I_1 - (1/r_{out})V_2 \tag{2.18a}$$

$$I_1 = (V_S - \alpha V_2)/(r_{in} + R_S). \tag{2.18b}$$

In this case, efficient coupling from source to load demands that r_{in} is small (I_1 large) and r_{out} large compared to R_L (I_2 large).

In Section 2.1, one simple method of finding R_{EQ} for Thevenin's equivalent circuit was to run down voltage souces to zero (i.e. short circuit them) and replace current sources by open circuits. This method fails if the circuit contains dependent sources. To illustrate this point, consider as an example the circuit of figure 2.19. On open circuit, $V_2 = V_1 - \alpha V_2$, so across terminals A and B we have $V_{EQ} = V_1/(1+\alpha)$. On short circuit, $I_{EQ} = V_1/R$. Hence $R_{EQ} = V_{EQ}/I_{EQ} = R_1/(1+\alpha)$. This is the correct result. If, however, one imagines the batteries shorted, ($V_2 = 0$), method (iii) of section 2.1 would lead to the incorrect result that $R_{EQ} = R_1$. The discrepancy arises because the circuit contains a dependent source and it is no longer permissible to scale voltages and currents down to zero.

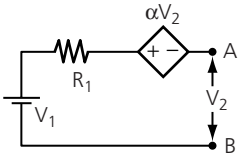


Fig. 2.19. Finding R_{EQ} with a dependent source.

Copyright © 2005 IOP Publishing Ltd.

2.8 Systems

Amplifiers are examples of building blocks which go into the construction of an electronic **system**. It is convenient to be able to break down the performance of a large system, such as that in figure 2.20, into sub-systems each of which performs a specific function. Each unit produces an output which is some function of its inputs, sometimes simply proportional to the input but sometimes involving differentiation or integration or perhaps a time delay. This function is called the **transfer function** T or T' of the unit. Each unit has inputs and outputs expressed in Thevenin or Norton equivalent form.

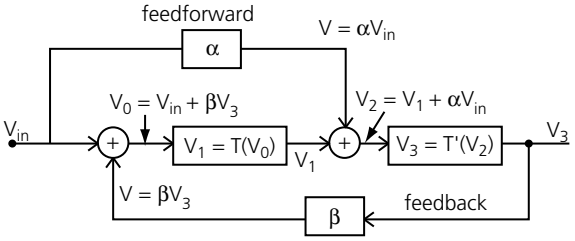


Fig. 2.20. An electronic system including feedback and feedforward.

Chapter 7 will demonstrate that the properties of the system can be radically altered by **feedback** of a signal to an early stage from a later one. For example, the output V_3 in figure 2.20 might be a signal controlling the speed of a motor; if V_{in} measures the required speed, it may be convenient to drive the electronics with a signal $(V_{in} - V_3)$ describing the discrepancy between required speed and actual speed. In this case, $\beta = -1$. This is an example of a servosystem. There are also situations where it is convenient to feed a signal forward, skipping intermediate stages; this is illustrated by αV_{in} in the figure.

The behaviour of figure 2.20 is described algebraically by a set of network equations involving the various voltages and transfer functions. Systems design concerns the analysis of these network equations, for example their time dependence and stability. Building a system is nowadays largely a matter of assembling units available commercially: power supplies, amplifiers, digital gates, analogue to digital converters and so forth. The following chapters will be concerned with understanding the performance possible from each type of unit.

2.9 Summary

When circuits are linear, Thevenin's theorem and Norton's offer short-cuts. These are very handy ways of simplifying and solving algebra of circuits. The Thevenin equivalent voltage across two terminals is the open circuit voltage, figure 2.2, and the Norton equivalent current source is equal to the short circuit current, figure 2.3.

Copyright © 2005 IOP Publishing Ltd.

For either, R_{EQ} is obtained by shorting batteries and replacing current sources by open circuits, figure 2.4. Amplifiers may be represented by Thevenin or Norton input and output circuits, with the output generator proportional to input current or voltage.

2.10 Exercises

1. *Thevenin's theorem.* A circuit produces voltage V on open circuit. When it is measured with a voltmeter having an impedance of 20 kΩ, the voltmeter reads 15 V; when the voltmeter is switched to a range having an impedance of 50 kΩ, it reads 18 V. What is V? (Ans: 20.8 V.)

2. *Thevenin's theorem.* (QMC). Calculate the currents I_1 and I_2 in the circuit of figure 2.21. Calculate also the potential difference across the current source and indicate its sign. (Ans: $I_1 = -0.6$ A, $I_2 = 0.4$ A, 32 V.)

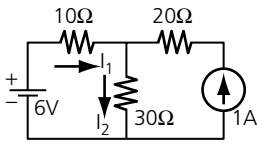


Fig. 2.21.

3. *Thevenin and Norton.* Find the Thevenin and Norton circuits equivalent to figure 2.22. (Ans: $V_{EQ} = V R_3/(R_1 + R_3)$; $R_{EQ} = (R_1R_2 + R_2R_3 + R_3R_1)/(R_1 + R_3)$; $I_{EQ} = V R_3/(R_1R_2 + R_2R_3 + R_3R_1)$.)

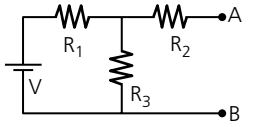


Fig. 2.22.

4. *Thevenin, Norton and superposition.* Find the voltage V across the load resistor R_L in figure 2.23. Find the Thevenin equivalent of the circuit supplying terminals AB and hence check V. (Ans: $V = 8R_L/\{3(4 + R_L)\}$; $V_{EQ} = 8/3$ V, $R_{EQ} = 4\ \Omega$; $I_{EQ} = 2/3$ A.)

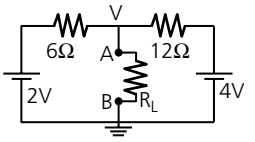


Fig. 2.23.

Copyright © 2005 IOP Publishing Ltd.

5. *Norton's theorem.* (RHBNC). Find the Norton equivalent circuit across terminals *AB* of figure 2.24. (Ans: $I_{EQ} = 0.75$ A, $R_{EQ} = 4\ \Omega$.)

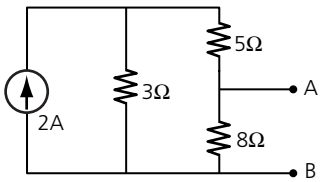


Fig. 2.24.

6**. *Series and parallel resistors, mesh currents, node voltages, Thevenin and Norton.* (a) Write down the equations for the mesh currents I_{1-4} in figure 2.25, but do not solve them. (b) Find the current $I_3 + I_4$ supplied by the battery by simplifying series and parallel combinations of resistors. (c) Find V_1 and V_2. (d) Hence find I_{1-4} and check them against (a). (e) Find the Thevenin equivalent of the circuit to the left of *AB*; use this to check I_4. (f) Find the Norton equivalent of the circuit across the terminals *CD*. Use this to check $I_3 + I_4$. (g) Find the Norton equivalent of the circuit to the right of *EF*; use it to check I_1. (Ans: (a) $I_3 + I_4 = 2$ A; (c) $V_1 = 8$ V; $V_2 = 8/3$ V; (d) $I_1 = 1/9$ A; $I_2 = 2/9$ A; $I_3 = 2/3$ A; $I_4 = 4/3$ A; (e) $V_{EQ} = 144/14$ V, $R_{EQ} = 24/14\ \Omega$; (f) $I_{EQ} = 3$ A, $R_{EQ} = 4\ \Omega$; (g) $I_{EQ} = 18/19$ A, $R_{EQ} = 24 \times 19/143\ \Omega$.)

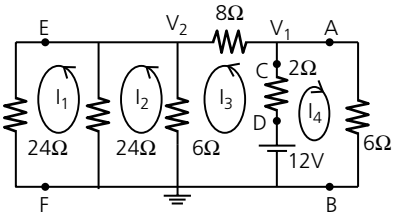


Fig. 2.25.

7. *Thevenin and Norton.* Obtain the Thevenin and Norton circuits equivalent to figure 2.26, first at output terminals *AB* then *AC*. (Ans: For *AB*,

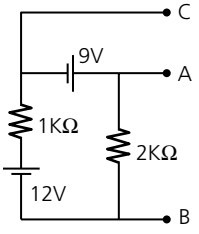


Fig. 2.26.

Copyright © 2005 IOP Publishing Ltd.

$V_{EQ} = 14$ V, $R_{EQ} = 2/3$ kΩ, $I_{EQ} = 21$ mA; for AC, $V_{EQ} = 9$ V, $R_{EQ} = 0$, $I_{EQ} = \infty$.)

8. *Sensitivity of a bridge.* (QMC). In the bridge circuit shown in figure 2.27, the bridge wire AB is 1 m long and has a resistance of 200 Ω; the galvanometer has a resistance of 37.5 Ω and gives a noticeable reading for a current of 10 μA. Construct an equivalent circuit for the voltage or current applied to the galvanometer. Hence find (i) the position of the sliding contact for balance; (ii) the distance the contact may be moved before the galvanometer gives a noticeable reading; here, take R_{EQ} at the balance points as an appoximation. (Ans: 62.5 cm from A; 0.625 mm.)

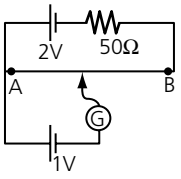


Fig. 2.27.

9. *Thevenin, Norton and nodes.* (QMC). Find the Thevenin and Norton equivalents of figure 2.28. (Ans: $V_{EQ} = 4/11$ V; $R_{EQ} = 10/11$ Ω; $I_{EQ} = 2/5$ A.)

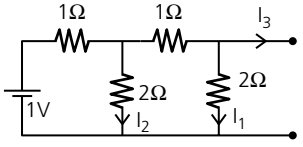


Fig. 2.28.

10*. *Thevenin.* (QMC). What is the Thevenin equivalent of the circuit in figure 2.29? (Ans: $V_{EQ} = 0$; $R_{EQ} = 4R/3$.)

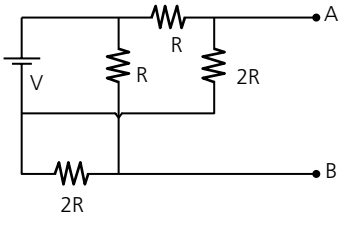


Fig. 2.29.

Copyright © 2005 IOP Publishing Ltd.

11*. *Thevenin.* (UC). Show that the Wheatstone bridge of figure 2.30 has an output voltage

$$V_{out} = \frac{E(R_2R_3 - R_1R_4)}{(R_1 + R_2)(R_3 + R_4)}.$$

If R_2 is a platinum resistance thermometer of resistance $R_t = R_0(1+\alpha t)$, where R_0 and R_t are the resistances at $0°C$ and $t°C$ respectively and R_1, R_3 and R_4 are all equal to R_0, show that V_{out} is proportional to t if $\alpha\ t \ll 1$. What is then the output impedance of the circuit, assuming the battery has negligible internal resistance? What must be the input impedance of a *DC* amplifier connected across *AB* to ensure the maximum power transfer to the *DC* amplifier from the thermometer circuit? (Ans: $V_{out} = \alpha Et/4$; R_0.)

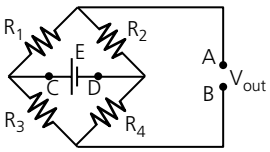


Fig. 2.30.

12. *Dependent current.* Find I and V_X in figure 2.31. What power is supplied by the battery and by the dependent current source? What powers are dissipated in the 3 kΩ, 2 kΩ and 6 kΩ resistors? (Ans: 0.4 mA, 4.8 V; 2.4 mW, 33.6 mW, 0.48 mW, 11.52 mW, 24 mW.)

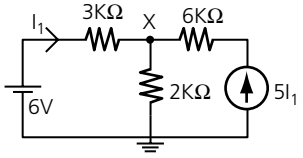


Fig. 2.31.

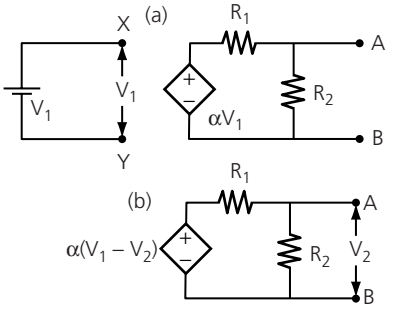


Fig. 2.32.

13. *Thevenin and dependent voltage.* In figure 2.32, terminals XY are open circuit. Find the Thevenin and Norton equivalents of the circuits to the left of AB. Note that in (b), R_{EQ} cannot be found by shorting voltage sources, because of the dependent voltages. (Ans: (a) $V_{EQ} = \alpha V_1R_2/(R_1 + R_2)$, $I_{EQ} = \alpha V_1/R_1$, $R_{EQ} = R_1R_2/(R_1 + R_2)$;

Copyright © 2005 IOP Publishing Ltd.

(b) $V_{EQ} = \alpha V_1 R_2/[R_1+R_2(1+\alpha)]$, $I_{EQ} = \alpha V_1/R_1$, $R_{EQ} = R_1 R_2/[R_1+R_2(1+\alpha)]$.)

14. *Thevenin and dependent current.* Find I_{EQ} then V_{EQ} hence R_{EQ} for the circuit feeding terminals AB of figure 2.33. For what value of α is $R_{EQ} < 0$? What value of α would you choose for an ideal current source? (Ans: $I_{EQ} = (14-2\alpha)5/84$ mA, $V_{EQ} = (70-10\alpha)/(20-2\alpha)$ V, $R_{EQ} = 42/(10-\alpha)$ kΩ; $\alpha > 10$; $\alpha = 10$.)

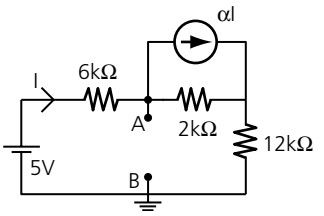


Fig. 2.33.

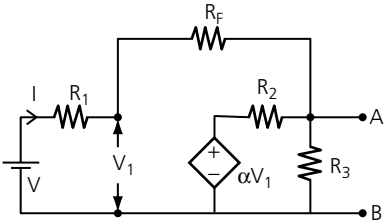


Fig. 2.34.

15*. *Thevenin and Norton.* Find I_{EQ} then V_{EQ} for the circuit to the left of terminals AB in figure 2.34. (Ans: $I_{EQ} = V(1+\alpha R_F/R_2)/(R_1+R_F)$; $V_{EQ} = V R_3(R_2+\alpha R_F)/[(R_1+R_F)R_2+R_3(R_1+R_F+R_2-\alpha R_1)]$.)

Copyright © 2005 IOP Publishing Ltd.

3

Capacitance

3.1 Charge and Capacitance

Capacitance plays a crucial role in understanding the speed at which electronics can switch. This is a fundamental issue in high speed computers and telecommunications. In a typical computer today, transistors switch in a nanosecond (10^{-9} s). Oscilloscopes capable of sampling and displaying waveforms at 10^{-10} s intervals are available (at a price). Oscillators working up to a frequency of 100 GHz (10^{11} cycles/s) are available. It turns out that capacitance is the feature which sets these limits.

The concept of capacitance will be introduced with an idealised example. Suppose an isolated plate is suspended in space. If charge is to be assembled on the plate, work must be done against the electrostatic repulsion between the charges. The necessary potential energy may be supplied by a battery, as in figure 3.1. If a second plate is connected to Earth and brought up close to the first one, as in figure 3.2, charges will flow from earth, producing a negative charge on the lower plate. The charges get as close to neutralising one another as the geometry allows, and ideally the charge on the lower plate exactly balances that on the first. In this case, the result is a system having net zero charge, with no external electric field. Viewed from a distance, it looks neutral (or more precisely, like a dipole). If the separation of the plates is s, there is an electric field $\mathbf{E} = V/s$ running directly from one plate to the other (figure 3.3).

This pair of plates is the simplest version of a **capacitor**: a system with no net charge, but balancing positive and negative charges stored on two separate plates. The charge Q stored on each plate is related to the applied voltage V by

$$\boxed{Q = CV} \tag{3.1}$$

where C is the **capacitance**, measured in **farads** (F). Capacitors can be made from plates of any shape. However, the linear relation between Q and V continues to hold, and can be proved in general from Coulomb's law. The magnitude of the capacitance depends on the geometry. For parallel plates of area A and separation s,

Copyright © 2005 IOP Publishing Ltd.

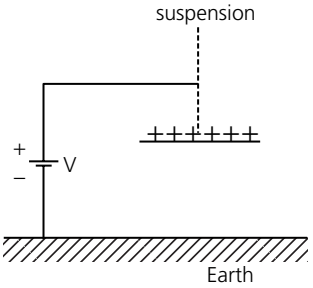


Fig. 3.1. Charging a plate.

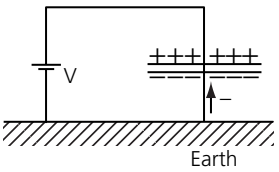


Fig. 3.2. A parallel plate capacitor.

$$C = \epsilon_0 A/s \tag{3.2}$$

where ϵ_0 is a fundamental constant $10^7/4\pi c^2 = 8.854 \times 10^{-12}$ F/m, and c is the velocity of light. The units are such that the capacitance C is usually very small. For example, if $A = 1$ cm^2 and $s = 0.1$ cm, $C = 0.885 \times 10^{-12}$ F. Thus capacitances are commonly from picofarads (10^{-12} F, written pF or $\mu\mu$F) to microfarads 10^{-6} F, written μF.

Switching speed

When a transistor switches, the voltages between its terminals change and this requires movement of charges. From equation (3.1),

$$\boxed{I = \frac{\mathrm{d}Q}{\mathrm{d}t} = C\frac{\mathrm{d}V}{\mathrm{d}t}} \tag{3.1a}$$

so the rate at which the voltage can change is governed by C and the available current I. Fast switching requires large I (hence small circuit resistance) or small C or both. One of the virtues of miniaturising transistors in integrated circuits is the reduction in capacitance: capacitances of 10^{-15} F/(μm)^2 are achieved.

The linear relation (3.1) between Q and V may be checked experimentally with the circuit of figure 3.4. When the switch is closed, the oscilloscope records a signal proportional to current I like that shown in figure 3.5. It is easy to demonstrate on the oscilloscope that the magnitude of I scales with V, but that otherwise the

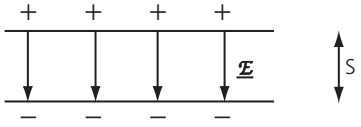


Fig. 3.3. Electric field within a capacitor.

Copyright © 2005 IOP Publishing Ltd.

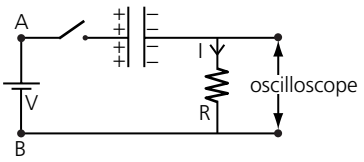


Fig. 3.4. Circuit to charge a capacitor.

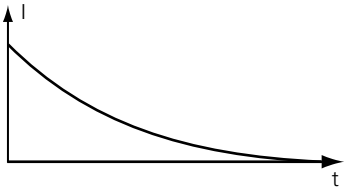


Fig. 3.5. Current v. time.

shape does not change. A formula for the curve will be derived later. The area under it measures the charge Q which has accumulated on each plate:

$$Q = \int_0^t I \, \mathrm{d}t. \tag{3.3}$$

When current flows 'through' the capacitor, positive and negative charges accumulate at the same rate on the two plates, so the current flowing into one plate equals that leaving the other, and there appears to be a continuous current through the capacitor.

The reality of the charge stored on the plates may be demonstrated by removing the battery and then shorting AB: the oscilloscope records a current in the reverse sense while the capacitor discharges. Whenever large DC voltages are applied, capacitors should be discharged after use, since the stored charge can be hazardous to anyone touching the terminals.

3.2 Energy Stored in a Capacitor

Energy must be supplied in order to assemble charge on the plates of a capacitor. This energy is given by

$$\begin{aligned} E &= \int IV \, \mathrm{d}t = \int \frac{\mathrm{d}Q'}{\mathrm{d}t} \frac{Q'}{C} \, \mathrm{d}t = \frac{1}{C} \int_0^Q Q' \, \mathrm{d}Q' \\ &= \tfrac{1}{2} Q^2/C = \tfrac{1}{2} QV = \tfrac{1}{2} CV^2. \end{aligned} \tag{3.4}$$

It is stored as potential energy and can be returned when the capacitor is discharged.

3.3 The Effect of a Dielectric

Suppose an isolated capacitor carries charge Q and voltage V and suppose an insulator is inserted between the plates. The electric field polarises the dielectric (figure 3.6) and the net field within the material is now equal to the original field less that due to the induced charges. Because the field drops, the potential difference between the plates likewise drops: remember $\mathbf{E} = V/s$. The charge Q has not

Copyright © 2005 IOP Publishing Ltd.

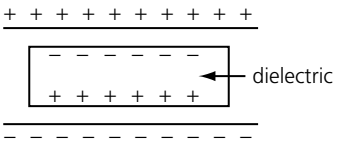


Fig. 3.6. Effect of a dielectric.

changed. Hence $C = Q/V$ has increased. It goes up by a factor ϵ which is called the **dielectric constant** of the material:

$$C \to \epsilon\epsilon_0 A/s. \tag{3.2a}$$

Alternatively, if the plates are connected to an external battery so as to keep V constant, then Q increases: more charge is stored.

3.4 Capacitors in Parallel

In the parallel arrangement shown in figure 3.7, there is the same voltage V across both capacitors. Thus

$$Q_1 = C_1 V \qquad Q_2 = C_2 V$$

and the total charge stored Q is

$$Q = Q_1 + Q_2 = (C_1 + C_2)V.$$

The arrangement behaves like an equivalent capacitor with capacitance

$$\boxed{C_{EQ} = C_1 + C_2.} \tag{3.5}$$

It is physically obvious that parallel capacitances will add in this way, because of the increase in the area on which charge is stored.

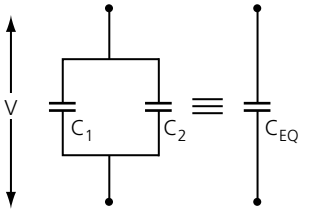


Fig. 3.7. Capacitors in parallel.

Copyright © 2005 IOP Publishing Ltd.

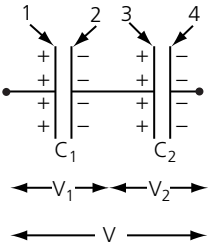


Fig. 3.8. Capacitors in series.

3.5 Capacitors in Series

In this case, figure 3.8

$$V = V_1 + V_2 = Q_1/C_1 + Q_2/C_2.$$

Because the plates labelled 2 and 3 are isolated electrically from the outside world, the net charge stored on them must be zero, so $Q_1 = Q_2$. The equivalent capacitance C_{EQ} is given by

$$V = Q/C_{EQ} = Q/C_1 + Q/C_2,$$

$$\boxed{\frac{1}{C_{EQ}} = \frac{1}{C_1} + \frac{1}{C_2}.} \tag{3.6}$$

Can you make sense of equations (3.5) and (3.6) in terms of the total energy stored, using equations (3.4)?

3.6 The *CR* Transient

The remainder of the chapter will be concerned with the question of switching speed. If, as in figure 3.9, a pulse generator applies a square pulse to a resistor and capacitor in series, the waveforms observed across C and R are as sketched in the figure. They are called **transients**. The voltage V_R across the resistor decays exponentially from an initial value V_0 according to

$$V_R \propto V_0 \mathrm{e}^{-t/CR}.$$

First we derive the exponential shape of V_R, then the result will be extended to more elaborate situations where steps of voltage follow one another faster than V_R can settle to zero.

Copyright © 2005 IOP Publishing Ltd.

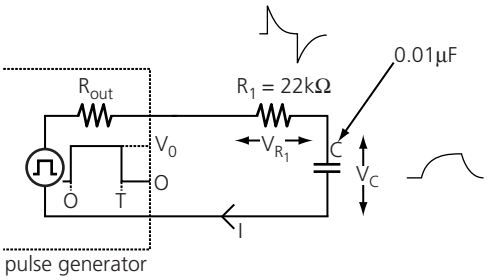


Fig. 3.9. Charging a capacitor.

In figure 3.9, suppose the capacitor is initially uncharged and the applied voltage is zero. At time $t = 0$, the generator applies a single square voltage pulse of height V_0 for a time T and then returns to zero. Physically what happens is that the capacitor charges towards the DC voltage V_0. Initially the capacitor is uncharged and all of V_0 is across $(R_{out} + R_1)$; current $I_0 = V_0/R$ flows, where R is the total resistance $(R_{out} + R_1)$. This current begins to charge the capacitor; the voltage across R then falls, and with it the current. Thus $I(t)$ falls steadily until either the capacitor is fully charged or the input pulse ends. When the latter happens, the sequence is reversed and the capacitor discharges through R_{out} and R_1.

Algebraically, we need to apply Kirchhoff's voltage law to the circuit. The voltage across the resistors is $V_R = IR$ and that across the capacitor is Q/C, so

$$V_0 = RI + \frac{Q}{C} = RI + \frac{1}{C}\int_0^t I(t')\,\mathrm{d}t'. \tag{3.7}$$

Differentiating with respect to time,

$$0 = R\mathrm{d}I/\mathrm{d}t + I/C. \tag{3.8}$$

This is a simple **differential equation**. It involves differentials only as high as $\mathrm{d}I/\mathrm{d}t$, and is therefore called a first-order differential equation. A circuit obeying such an equation is called a **first-order system**.

Exponential decay

From (3.8),

$$\mathrm{d}I/\mathrm{d}t = -I/CR. \tag{3.8a}$$

Current decreases at a rate proportional to its magnitude. The curve of figure 3.5 clearly has this property. This type of equation arises very frequently and its solution is the **exponential** function. Its properties are reviewed systematically in Appendix B. It occurs so commonly (like trig functions) that it is available on most hand calculators.

Copyright © 2005 IOP Publishing Ltd.

The equation is solved by writing

$$\mathrm{d}I/I = -\mathrm{d}t/CR$$

$$\ln_{\mathrm{e}} I = -t/CR + \text{constant}$$

$$\boxed{I = I_0 \exp(-t/CR).}$$

The solution may be checked by differentiating it and substituting back into equation (3.8a):

$$\mathrm{d}I/\mathrm{d}t = -(1/CR)I_0 \exp(-t/CR) = -I/CR.$$

The differential equation does not determine I_0, which arises from the integration constant. Any first-order differential equation involves one such constant. To find it, the **initial conditions** must be considered. At time $t = 0$, $I = I_0 = V_0/R$, so

$$I = (V_0/R)\exp(-t/\tau).$$

Here

$$\boxed{\tau = CR} \tag{3.9}$$

is called the **time constant** (or relaxation time). It is the time in which I falls from I_0 to $I_0/\mathrm{e} \simeq 0.37I_0$. If C is in farads and R in ohms, τ is in seconds. As a numerical example, suppose $C = 0.01\ \mu\mathrm{F}$ and $R = 22\ \mathrm{k}\Omega$; then $\tau = 0.22$ ms. Finally, the algebraic expressions for all waveforms are:

$$\begin{aligned} V_R &= V_0\mathrm{e}^{-t/\tau} \\ V_{R1} &= (R_1/R)V_R \\ V_C &= V_0(1 - \mathrm{e}^{-t/\tau}) \\ Q &= CV_0(1 - \mathrm{e}^{-t/\tau}). \end{aligned} \tag{3.10}$$

The form of these results is displayed in figure 3.10 for several values of τ. This is known as the **transient response** of the circuit, i.e. its response to a change in DC conditions.

Show that the tangent to $V_R(t)$ at $t = 0$ goes through $V = 0$ at $t = \tau$.

It is well worth experimenting yourself with these waveforms using an oscilloscope and square waves from a pulse generator. Suppose you want to determine τ experimentally. Make the height of the square wave 2.7 cm on the oscilloscope trace. Then τ is given by the time at which V_C deviates from its final value by 1 cm; at this time, V_R will read 1 cm on the trace.

Copyright © 2005 IOP Publishing Ltd.

In the circuit of figure 3.9, the source supplies charge Q at voltage V_0 and hence delivers energy QV_0; half of this is stored in the capacitor, but where does the other half go?

All practical devices have some capacitance. An oscilloscope has an input capacitance of typically 10–25 pF. The coaxial cable used to connect to an oscilloscope has a capacitance of typically 150 pF per metre. If they are to follow an input signal rapidly, this capacitance must be charged quickly, i.e. V_C must follow a step in voltage as closely as possible. This requires small τ, i.e. small CR.

In a digital circuit, a capacitance of 10 pF *is to be charged to* 5 V *through a 50* Ω *resistor. What is* τ *and what is the initial charging current?*

If the input pulse is long enough, V_C eventually reaches V_0 and $V_R \rightarrow 0$. Suppose, however, that the input ends before the capacitor is fully charged, as in the lower half of figure 3.10. The subsequent discharge of the capacitor is governed by algebra very similar to the charging process, but with different initial conditions. We now find a set of expressions analogous to equation (3.10) describing the resulting waveforms.

The charge on the capacitor cannot change instantly when the input voltage drops to zero. The change V_0 in the input pulse must therefore appear in full in the waveform V_R. The new voltage across the resistor is $-V_1$ and current $-V_1/R$ flows initially. The differential equation (3.8) has not changed, so the capacitor discharges from voltage V_1 towards zero with the same time constant τ as before.

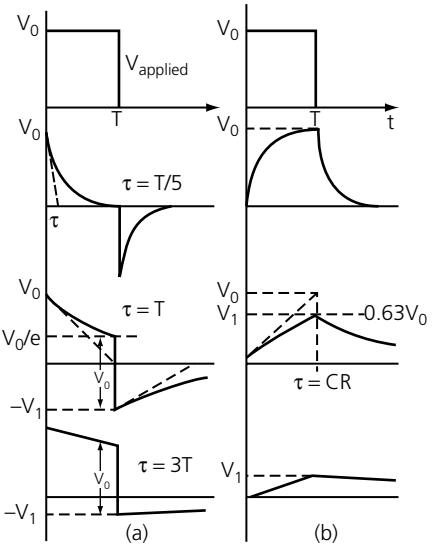


Fig. 3.10. (a) V_R and (b) V_C v. t.

Copyright © 2005 IOP Publishing Ltd.

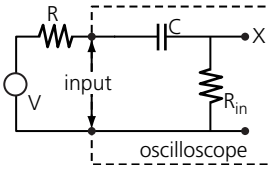


Fig. 3.11. AC input to an oscilloscope.

The previous expressions for I, V_R, etc. can be taken over with t replaced by $(t - T)$;

$$I = -(V_1/R)\mathrm{e}^{-(t-T)/\tau}$$

$$V_R = -V_1\mathrm{e}^{-(t-T)/\tau}$$

$$V_{R1} = (R_1/R)V_R$$

$$V_C = V_1\mathrm{e}^{-(t-T)/\tau}$$

$$Q = CV_C.$$

These results are used in figure 3.10. It is recommended that you get familiar with exponentials by plotting out a waveform for one value of T (e.g. $T = \tau$) using a hand calculator.

After a capacitor has charged up, the full voltage V_0 from the source appears across it and the voltage across the resistors is zero. The capacitor is said to block DC, because no current flows; this is one of its common uses. The input to an oscilloscope labelled 'AC' has such a capacitor, as shown in figure 3.11. The AC input is useful for eliminating unwanted DC levels. For example, you may want to look at a very small AC signal superimposed on a large DC background and it would be inconvenient if the scale of the display is dictated by the DC signal. The time constant of the circuit in figure 3.11 is $C(R + R_{in})$. The value of C is made large so that AC current flows as if the capacitor were transparent. In practice, CR_{in} is chosen to be ~ 0.1 s. The capacitor is called a **coupling capacitor** and the voltage source is said to be **AC coupled** to the oscilloscope.

Leakage resistance

No insulator is perfect, so in practice charge in an isolated capacitor will gradually leak through the dielectric and the capacitor will discharge. An equivalent circuit describing this is shown in figure 3.12. The time constant for the discharge is CR_{leak}.

Copyright © 2005 IOP Publishing Ltd.

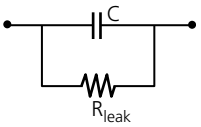


Fig. 3.12. Leakage resistance.

3.7 AC Coupling and Baseline Shift

Suppose the input signal to a *CR* circuit figure 3.13(a) has the irregular shape in figure 3.13(b). The voltage V_C across the capacitor follows with time constant *CR* as on figure 3.13(c).

Consider next what happens if the time constant *CR* is large and a train of square waves, as in figure 3.14(a), is applied from a pulse generator to the circuit of figure 3.11. The ratio p/q of the times for which the input is high and low respectively is called the **mark-to-space ratio**. Because of the large time constant, the circuit responds sluggishly and V_C tends towards the mean value of the input voltage,

$$\bar{V}_C = V_0 p/(p+q).$$

At the arrival of the first pulse, the capacitor begins to charge towards V_0, but it reaches only $V_1 \simeq V_0 t/\tau$ before the end of the pulse. The capacitor then begins discharging according to $V_C = V_1 \exp(-t/\tau)$. The discharge is incomplete before the second pulse arrives, so the capacitor acquires some net charge during the first cycle. This process continues during subsequent pulses, as in figure 3.14(b). Eventually equilibrium is reached when the rates of charging and discharging are equal.

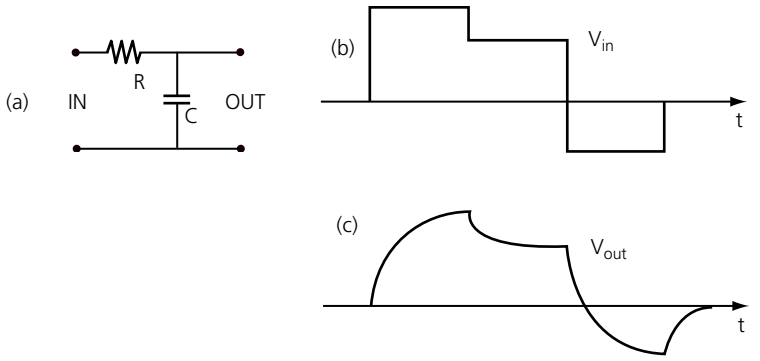


Fig. 3.13. Response of a CR circuit to an irregular waveform.

Copyright © 2005 IOP Publishing Ltd.

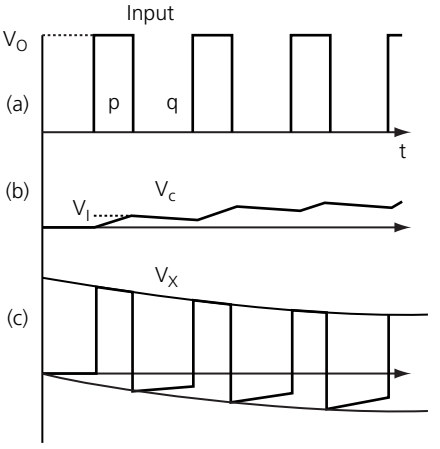


Fig. 3.14. Baseline shift generated by charging a capacitor.

It is easy to find $\overline{V}_c$. The charging current over the duration of the square pulse is $(V_0 - V_C)/R'$, where $R' = R_{in} + R$; the discharging current in the absence of the input pulse is V_C/R'. When these two currents are equal on average:

$$(V_0 - \bar{V}_C)p/R' = \bar{V}_C q/R'$$

$$\bar{V}_C = V_0 p/(p+q).$$

The voltage V_X across R_{in} drifts negative by $\bar{V}_C R_{in}/R'$, due to the discharge current flowing through the resistances, figure 3.14(c). This is called **base-line shift**. The discharge current may be put to use to make a simple **ratemeter**, producing an output signal proportional to the rate of the input pulses. However, in many situations it is unwelcome, since the voltage at X does not reproduce faithfully the input voltage, but has a DC shift superimposed. If, for example, the height of the pulses above zero plays a crucial role in subsequent circuitry, the DC shift can upset the intended operation. For high-speed circuitry, where pulses follow one another in quick succession, it is desirable to eliminate coupling capacitors by using DC coupling. Alternatively, the pulses may be shaped deliberately to have equal positive and negative areas, as in figure 3.15; such pulses are called **bipolar**.

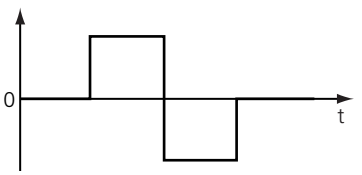


Fig. 3.15. Bipolar pulse.

Copyright © 2005 IOP Publishing Ltd.

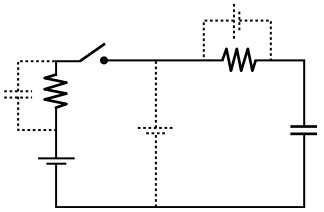


Fig. 3.16. Stray capacitance.

3.8 Stray Capacitance

In an electrical circuit, most of the electric field is along the wires and through components. However, a very small part of the electric field is through the air. In figure 3.1, lines of electric field run from charges on the plate to induced charges in the nearby earth and elsewhere. Associated with them is **stray capacitance**. Thus, to figure 3.9 one should strictly add the capacitances shown by the dashed lines on figure 3.16. Such stray capacitance is very difficult to calculate, but is of the order of magnitude $\epsilon_0 A/s$, where A is the area of an element and s is the separation of two elements, i.e. of order pF or less. Despite its small magnitude, stray capacitance ultimately limits high frequency performance.

Electrical circuits are usually mounted inside a chassis which is earthed; this eliminates stray capacitance to distant sources and hence capacitative pick-up. Likewise, signal cables are protected by an earth braid. Shielding and earthing is a complicated art, and the reader is referred to specialist discussions (e.g. Horowitz and Hill, 'The Art of Electronics', CUP).

3.9 Integration and Differentiation

The algebra of the previous sections applies specifically to the case when the applied voltage is constant or a square wave. More generally, for an applied voltage $V(t)$ with arbitrary time dependence,

$$V_C = \frac{1}{C}\int I\ \mathrm{d}t' = \frac{1}{CR}\int (V - V_C)\ \mathrm{d}t'$$

where R is the total series resistance in the circuit. For large values of CR, the capacitor charges slowly and V_C is small. In this case, if the capacitor is uncharged initially,

$$V_C \simeq \frac{1}{CR}\int_0^t V(t')\ \mathrm{d}t'. \tag{3.11}$$

For this reason, the circuit of figure 3.17(a) with large CR is called an **integrator**. If a pulse is applied to the circuit, the capacitor is said to **integrate** the input pulse. The snag is that the output signal is small, so CR should not be *too* large. The approximation in equation (3.11) is accurate only for $t \ll CR$. To see this, the

Copyright © 2005 IOP Publishing Ltd.

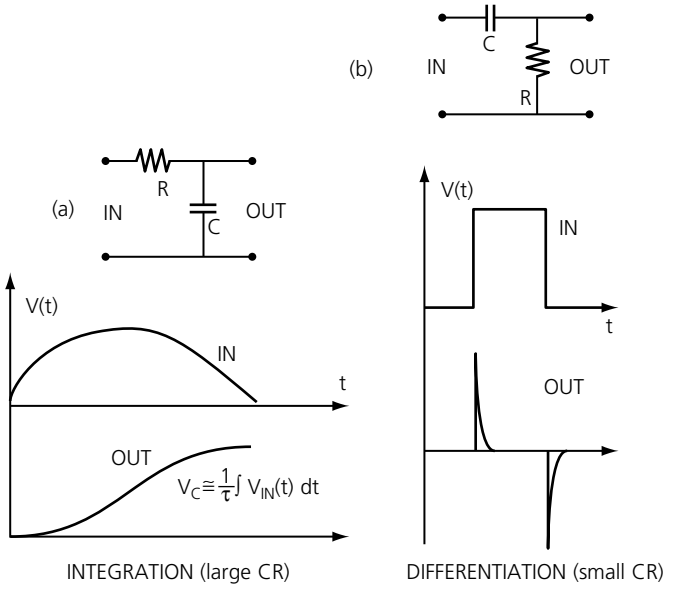


Fig. 3.17. (a) Integration, (b) Differentiation.

exponential in equation (3.10) is expanded in a power series with V equal to a constant V_0. The result is

$$V_C = V_0 \left\{ \frac{t}{CR} - \frac{1}{2}\left(\frac{t}{CR}\right)^2 + \ldots \right\}.$$

Equation (3.11) gives only the first term in the bracket.

The other extreme arises when CR is small. Figure 3.17(b) shows that in this case V_R displays a narrow spike at the leading and trailing edges of the input pulse. The output V_R is large where the slope of the input voltage is large. For this reason the circuit is known as a **differentiating** circuit and the process of generating the spike is called **differentiation**. This arrangement may be used to generate a spike which can be used as a trigger in subsequent circuits; using the negative spike on the trailing edge is a simple way of generating a **time delay** equal to the width of the square pulse.

3.10* Thevenin's Theorem Again

Suppose you want to measure the input capacitance of your oscilloscope (and its leads). This can be done by measuring the time constant for charging it. However, it is necessary to allow for the input resistance R_2 in parallel with C, figure 3.18.

Copyright © 2005 IOP Publishing Ltd.

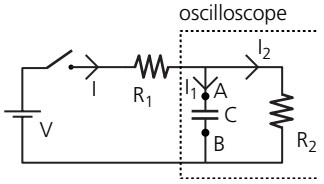


Fig. 3.18. Illustrating Thevenin's theorem.

The algebra goes much the same as before, except that it is necessary to account separately for I_1 and I_2:

$$I_2R_2 = V - IR_1 = V - (I_1 + I_2)R_1$$

$$\text{or} \qquad I_2 = (V - R_1I_1)/(R_1 + R_2).$$

Then

$$V_c = \frac{1}{C}\int I_1 \, \mathrm{d}t' = I_2R_2 = \frac{VR_2}{R_1 + R_2} - \frac{R_1R_2I_1}{R_1 + R_2}.$$

This is an equation for I_1 of precisely the same form as equation (3.7) except that V_0 has been replaced by $VR_2/(R_1+R_2)$ and R has been replaced by $R_1R_2/(R_1+R_2)$. Obviously, we have constructed a Thevenin equivalent for the circuit across AB, i.e. across the capacitor. With AB open circuit, the voltage across AB is reduced from V to $VR_2/(R_1 + R_2)$, i.e. that given by the potential divider made by R_1 and R_2; and the resistance of the Thevenin equivalent is given by R_1 and R_2 in parallel from A to B (with the battery short-circuited).

3.11 Summary

$$Q = CV$$

$$I = C\mathrm{d}V/\mathrm{d}t.$$

Parallel capacitors have capacitance $C_1 + C_2$; in series, $C = C_1C_2/(C_1 + C_2)$. The time constant for charging through resistance R is $\tau = CR$. For a square wave input, $V_C = V_0(1 - \mathrm{e}^{t/CR})$ and $V_R = V_0\mathrm{e}^{-t/CR}$. For large t, $V_C \to V_0$ and $V_R \to 0$. The voltage across a capacitor cannot change instantaneously.

3.12 Exercises

1. For a capacitor and resistor in series, why is the time constant τ small if either C or R is small? Why is it dangerous to risk touching a large capacitor?

Copyright © 2005 IOP Publishing Ltd.

2. What are the differences between a battery and a capacitor?

3. (Chelsea). Why is the Wheatstone bridge not used for the measurement of very high resistances? Describe how you would measure a resistance of the order of 10 MΩ by allowing a capacitor to discharge through it. Indicate suitable values for the components to be used. How can a check be made in order to discover whether the capacitor itself is 'leaky'?

4. A capacitor of 2 μF is connected across a 12 V battery. What charge and what energy are stored in the capacitor? (Ans: 24 μC, 0.144 mJ.)

5. The current through a 10 μF capacitor is $I = (10t - 4t^2)$ μA from $t = 0$ to 2.5s. If the capacitor is initially uncharged, what is the voltage across it as a function of time? (Ans: $0.1(5t^2 - 4t^3/3)$ V.)

6. The voltage across a 1.5 μF capacitor is $V = (10t - 4t^2)$ V from $t = 0$ to 2.5 s. What current flows through it as a function of time? (Ans: $1.5(10 - 8t)$ μA.)

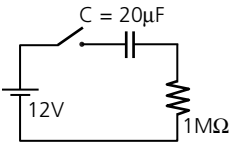


Fig. 3.19.

7. Initially the switch in figure 3.19 is open and the capacitor in uncharged. At time $t = 0$ it is closed. Find the voltage $V_C(t)$ across the capacitor against time t. (Ans: $12(1 - e^{-t/20})$ V, with t in s.)

8. *Thevenin.* At time $t = 0$, the capacitor of figure 3.20 carries no charge and the switch is moved to position A. After one time constant, it is moved to position B. Sketch the voltage V_R across the resistor as a function of time and label your sketch with quantitative values. (Ans: figure 3.21)

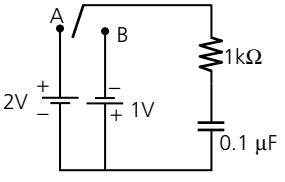


Fig. 3.20.

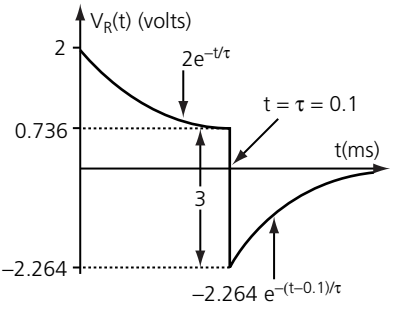


Fig. 3.21.

Copyright © 2005 IOP Publishing Ltd.

9. Initially the switch in figure 3.22 is open and the capacitor is uncharged. At time $t = 0$ the switch is closed. Find the voltage $V_C(t)$ across the capacitor as a function of time. (Ans: $4.125\{1 - \exp(-t/0.1375)\}$ V, with t in s.)

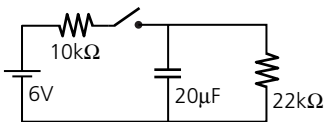


Fig. 3.22.

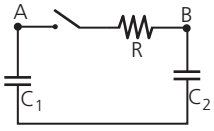


Fig. 3.23

10**. (QMC). In the circuit of figure 3.23, C_1 initially carries charge q. At time $t = 0$, the switch is closed. What is the current through R as a function of time? Show that equilibrium is reached eventually with voltage $q/(C_1 + C_2)$ across each capacitor and with signs such that no current flows through the resistor. How much energy is dissipated in the resistor after closing the switch? (Ans: $I = q(1 - e^{-t/C'R})/C_1R$ where $C' = C_1C_2/(C_1 + C_2)$; $E = \frac{1}{2}q^2C_2/\{C_1(C_1 + C_2)\}$.)

11*. Predict the output voltage waveform from the two circuits in figure 3.24 when the input voltage is a saw-tooth pulse with duration T. (Ans: figure 3.25)

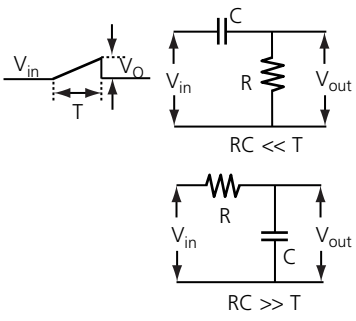


Fig. 3.24.

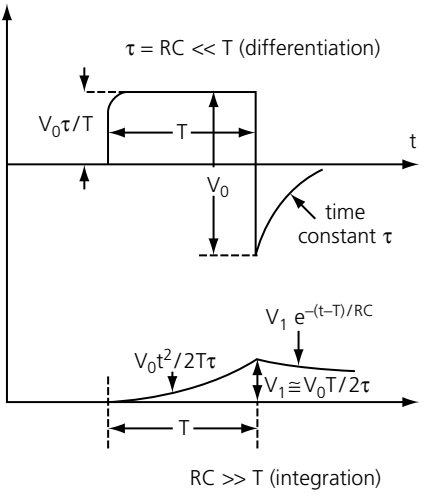


Fig. 3.25.

12. (QMC). In the circuit of figure 3.26, V_1 is a pulse generator giving square pulses which rise from zero to 6 V and last for 1 ms. Sketch the waveforms across AB when (a) a 5 kΩ resistor is applied there, (b) when a

Copyright © 2005 IOP Publishing Ltd.

0.1μF capacitor is connected there. Label your sketches with magnitudes describing the waveforms quantitatively. (Ans: figure 3.27)

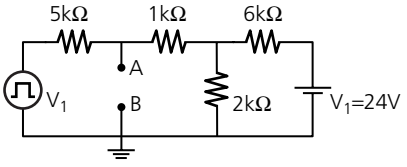


Fig. 3.26.

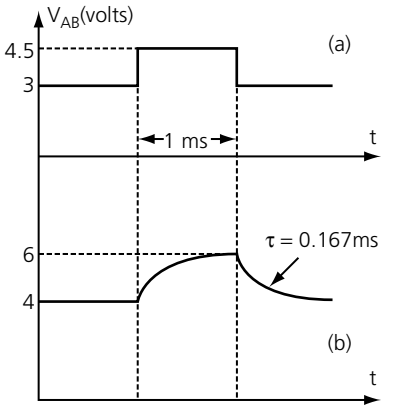


Fig. 3.27.

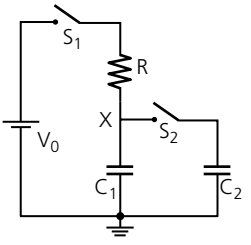


Fig. 3.28.

13*. (Chelsea). In figure 3.28, the capacitors C_1 and C_2 are both initially uncharged and keys S_1 and S_2 are open. S_1 is now closed, and after C_1 has charged fully, S_2 is also closed. Find the voltage across C_2 at time t after closing S_2. Outline in words how the behaviour is modified if a small resistance is in series with S_2. (Ans: C_1 charges to V_0 through R with $\tau = C_1R$. When S_2 is closed, C_1 and C_2 share charge and $V_X \to C_1V_0/(C_1+C_2)$. The parallel combination of C_1 and C_2 charges through R: $V_X = C_1V_0/(C_1 + C_2) + \{V_0C_2/C'\}\{1 - \exp(-t/RC')\}$, where $C' = C_1 + C_2$. Providing $R(S_2) \ll R$, C_1 and C_2 come to equilibrium with $\tau = R(S_2)C_1C_2/(C_1 + C_2)$; then as before.)

Copyright © 2005 IOP Publishing Ltd.

4

Alternating Current (AC); Bandwidth

4.1 Introduction

In the previous chapter we studied the charging of a capacitor by a voltage step. Now we turn to the way it responds to voltage $V_0 \cos \omega t$ or $V_0 \sin \omega t$. This is an **alternating voltage**. An example might be a radio signal. The voltage supplied by the mains is also close to a sine wave. Later, Chapter 15 will show that *any* voltage waveform $V(t)$ can be expressed by Fourier's theorem as a sum of sines and cosines. So if we can understand how a circuit responds to a sine wave, we can, with some effort, reconstruct its response to an input of any time dependence.

In Britain the mains voltage has a peak value $V_0 \simeq 240\sqrt{2}$ volts; in the rest of Europe it is $220\sqrt{2}$ volts and in the USA $110\sqrt{2}$ volts. The factor $\sqrt{2}$ will be explained shortly. One cycle $V_0 \cos \omega$t takes a time T such that $\omega T = 2\pi$, figure 4.1(a); so the time period $T = 2\pi/\omega$. The frequency is $f = 1/T = \omega/2\pi$. In Europe, $f = 50$ Hertz (or cycles/second); in the USA, $f = 60$ Hertz.

4.2 Power in a Resistor: RMS Quantities

The origin of the factor $\sqrt{2}$ will now be examined. The power dissipated in a resistor is $P = VI$. If a voltage $V = V_0 \sin \omega t$ (figure 4.1(a)) is applied directly across resistance R, the current $I = V_0 \sin \omega t/R$ and the instantaneous power is

$$P = VI = (V_0^2/R) \sin^2 \omega t = (V_0^2/2R)(1 - \cos 2\omega t).$$

Integrated over a complete cycle, the term $\cos 2\omega t$ is zero, so the mean power dissipated is $\bar{P} = V_0^2/2R$; the mean value of V^2 over the cycle is $\frac{1}{2}V_0^2$. Figure 4.1(b) demonstrates that V^2 is symmetrical about $\frac{1}{2}V_0^2$ if you turn the page upside down. To bring the expression for $\bar{P}$ into line with the DC result $P = V^2/R$, it is conventional to rewrite this expression as $\bar{P} = V_{RMS}^2/R$, where

Copyright © 2005 IOP Publishing Ltd.

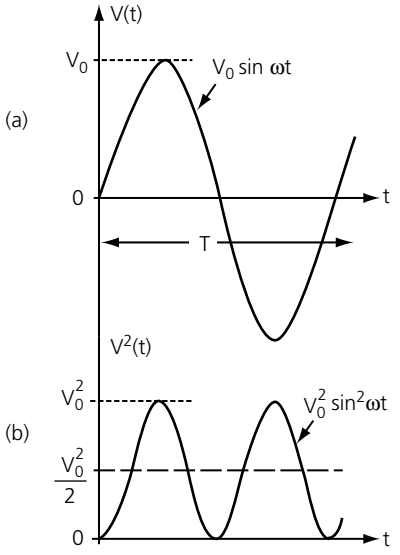


Fig. 4.1. (a) $V_0 \sin \omega t$ and (b) $V_0^2 \sin^2 \omega t$.

$$\boxed{V_{RMS} = V_0/\sqrt{2}} \tag{4.1}$$

RMS stands for root mean square and indicates that the square of V_{RMS} gives the mean value of $V^2(t)$. You may verify the factor $\sqrt{2}$ by measuring V_{RMS} of an AC signal on the AC range of a multimeter and determining V_0 by displaying the signal on an oscilloscope.

The mains are used largely to provide heating and lighting and it is conventional to refer to V_{RMS} rather than V_0: in Britain, $V_{RMS} = V_0 \cos \omega t$ with $V_0 = 240\sqrt{2}$ V. Likewise, $I_{RMS} = I_0/\sqrt{2}$ where $I = I_0 \cos \omega t$. So

$$\boxed{\bar{P} = V_{RMS} I_{RMS} = V_{RMS}^2/R = I_{RMS}^2 R.} \tag{4.2}$$

4.3 Phase Relations

For a resistance, $V_R = IR$; so an alternating current produces an alternating voltage and vice versa. For a capacitance, $V_C = (I/C) \int I \, dt$. The integral of a sine is a cosine, so an alternating current again leads to an alternating voltage and vice versa. However, the change from sine to cosine implies a shift in time dependence between current and voltage. Later, we shall find the same to be true for an inductor too.

Copyright © 2005 IOP Publishing Ltd.

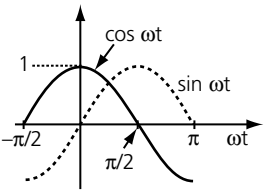


Fig. 4.2. Relations between sines and cosines.

The quantity ωt is called the **phase** of the voltage $V_0 \sin \omega t$. We could equally well have written $V = V_0 \cos \omega t$, or more generally $V = V_0 \cos(\omega t + \phi)$ by choosing the origin of time, $t = 0$, differently. Results are independent of which form is chosen. However, the phase *relations* amongst different quantities will be important, and it will be necessary to manipulate expressions like $\cos(\omega t + \phi)$. Figure 4.2 provides a brief reminder of some of the simple properties of sines and cosines:

$$\cos 0 = 1 \qquad \sin 0 = 0$$
$$\cos \pi/2 = 0 \qquad \sin \pi/2 = 1$$
$$\cos(-\phi) = \cos \phi$$
$$\sin(-\phi) = -\sin \phi$$
$$\cos(\omega t - \pi/2) = \sin \omega t$$
$$\cos(\omega t + \pi/2) = -\sin \omega t.$$

The last two are special cases of the important general formulae:

$$\cos(\omega t - \phi) = \cos \omega t \cos \phi + \sin \omega t \sin \phi$$
$$\cos(\omega t + \phi) = \cos \omega t \cos \phi - \sin \omega t \sin \phi.$$

If you have trouble remembering these expressions (particularly the signs), try trial substitutions with $\omega t = 0$ and $\pi/2$.

It is often convenient to represent an AC voltage V graphically as a point, figure 4.3(a), which describes a circle with angular frequency ω. The line OV

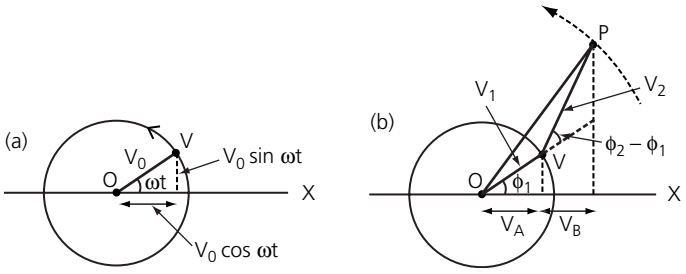


Fig. 4.3. (a) Vector representation of an AC voltage, (b) addition of two alternating voltages.

Copyright © 2005 IOP Publishing Ltd.

makes an angle ωt with OX and the length of OV is V_0. Thus V is described by a rotating vector like the beam of a lighthouse. Such a diagram is called a **phasor diagram** and the vectors are called **phasors**. Later it will emerge that they are in fact representations of complex numbers. The projection of OV on to the horizontal axis is $V_0 \cos \omega t$ and that on to the vertical axis is $V_0 \sin \omega t$, so this diagram can describe either form, sines or cosines. The angular frequency ω is measured in radians/second. The time period T is the time required for the point to move once round the circle.

An an illustration of the use of this representation, consider the addition of two voltages:

$$V_A = V_1 \cos(\omega t + \phi_1) \qquad V_B = V_2 \cos(\omega t + \phi_2).$$

Their sum is represented by the point P on figure 4.3(b), which has projection $V_A + V_B$ on to OX. The angle PVO is $\pi - (\phi_2 - \phi_1)$, so

$$OP = \{V_1^2 + V_2^2 + 2V_1V_2 \cos(\phi_2 - \phi_1)\}^{1/2}.$$

This result is independent of t, so the sum is also a voltage of angular frequency ω. The sum P is said to *lead* V_A because its phase angle is larger than ϕ_1 and the vector OP is describing its circle in advance of OV. Conversely, V_A is said to *lag* P and V_B.

4.4 Response of a Capacitor to AC

Now consider the relation between AC voltage and current for a capacitor alone, figure 4.4. If a current $I_0 \cos \omega t$ flows, charge on the capacitor and the voltage across it are given by

$$Q = \int I \, \mathrm{d}t$$

$$V = \frac{1}{C} \int I_0 \cos \omega t \, \mathrm{d}t = \frac{I_0}{\omega C} \sin \omega t$$

$$= \frac{I_0}{\omega C} \cos(\omega t - \pi/2). \tag{4.3}$$

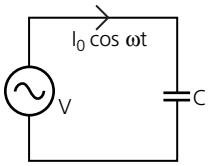


Fig. 4.4. Alternating voltage applied to a capacitor.

Copyright © 2005 IOP Publishing Ltd.

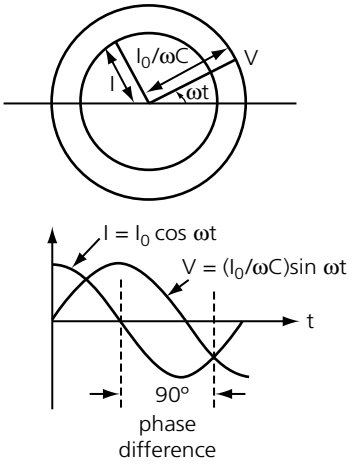


Fig. 4.5. The phase relation between voltage and current for a capacitor.

The magnitude of the voltage is given by $I_0/(\omega C)$. The quantity $1/\omega C$ plays a role similar to resistance in Ohm's law. It is called the **reactance** of the capacitor. The reactance drops to zero as $\omega \to \infty$; at high frequency, a capacitor acts like a short circuit. Conversely, as $\omega \to 0$, the current drops to zero: a capacitor blocks DC.

However, despite the similarity to Ohm's law, there is an essential difference: I leads V in phase by 90° or $\pi/2$, as shown on figure 4.5. Current must flow first if charge, hence voltage, is to become established on the capacitor.

In a capacitor, voltage follows current as charge reaches the plates. Does this suggest a physical explanation of why the reactance $\to 0$ *as* $\omega \to \infty$*?*

An AC voltage is applied to capacitors C_1 *and* C_2 *in parallel. Show that they behave like a single capacitor* $(C_1 + C_2)$.

4.5 Simple Filter Circuits

The remainder of this chapter focusses on the frequency response of more elaborate circuits containing a capacitor; this introduces the fundamental notion of **bandwidth**. Two circuits which occur very commonly are shown in figure 4.6. In outline, their behaviour follows from the fact that the reactance of the capacitor drops to zero as $\omega \to \infty$. A voltage will appear across the capacitor only at low frequencies; at high frequencies it is entirely across the resistor. Thus figure 4.6(a) transmits an output only at low frequencies and is called a **low pass filter**. Figure 4.6(b) does the converse; it is a **high pass filter**.

Suppose the AC voltage is switched on and left running long enough for transients to die away. This gives the **steady state** behaviour. Also assume no load

Copyright © 2005 IOP Publishing Ltd.

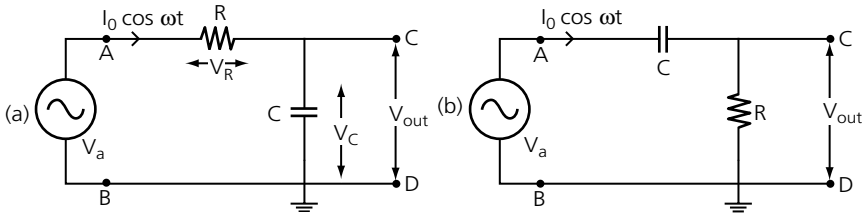


Fig. 4.6. (a) Low-pass filter, (b) high-pass filter.

is applied externally to terminals CD. If the current flowing round the circuit is $I_0 \cos \omega t$, the applied voltage is

$$\begin{aligned} V_a &= V_R + V_C \\ &= IR + (1/C) \int I \, \mathrm{d}t \\ &= RI_0 \cos \omega t + (1/\omega C) I_0 \sin \omega t. \end{aligned} \tag{4.4}$$

We anticipate that the voltage V_a lags behind current by an angle between $\pi/2$ (the result for a capacitor alone) and 0 (the result for a resistor). So we write

$$V_a = V_0 \cos(\omega t - \phi) = V_0 \{\cos \omega t \cos \phi + \sin \omega t \sin \phi\}.$$

If the coefficients of $\cos \omega t$ and $\sin \omega t$ are to agree with (4.4):

$$V_0 \cos \phi = RI_0 \tag{4.5}$$

$$V_0 \sin \phi = I_0/\omega C. \tag{4.6}$$

Squaring these equations and adding,

$$V_0^2 = \{R^2 + (1/\omega C)^2\} I_0^2; \tag{4.7}$$

dividing (4.6) by (4.5)

$$\tan \phi = 1/\omega CR. \tag{4.8}$$

These results justify the trial solution $V_a = V_0 \cos(\omega t - \phi)$.

Phasors

It is convenient to draw a snapshot at $t = 0$. The vector describing the current is then along the horizontal axis and so is the vector describing V_R. The voltage V_C lags I in phase by 90°. The vectors OA and OB add to give OC, representing the applied voltage V_a. We could have derived V_a graphically like this, bypassing the algebra.

Copyright © 2005 IOP Publishing Ltd.

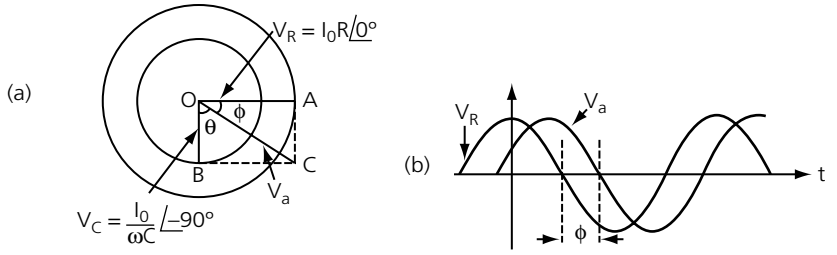


Fig. 4.7. (a) Phasor diagram for the CR filter circuit, (b) waveforms.

The magnitude of V_a is given by the hypotenuse of the triangle, i.e. by equation (4.7). The output of the high pass filter, V_R, leads V_a by angle ϕ, figure 4.7(b). When R dominates (large ω), $\phi \to 0$; when C dominates (at low frequencies), $\phi \to 90°$. The output of the low pass filter, V_C, lags V_a by angle θ, where

$$\tan\theta = \omega CR. \tag{4.9}$$

The magnitudes of V_R and V_C are given from equation (4.7) by:

$$|V_R| = I_0R = V_0R/\{R^2 + (1/\omega C)^2\}^{1/2} = V_0/\{1 + (1/\omega CR)^2\}^{1/2} \tag{4.10}$$

$$|V_C| = I_0/\omega C = V_0/\{1 + (\omega CR)^2\}^{1/2}. \tag{4.11}$$

When it comes to doing problems, the phasor diagram holds the key. The first step is to sketch this phasor diagram, putting in the angles θ and ϕ. For the series CR circuit we are considering, the voltage across the resistor is smaller than the applied voltage V_a by the factor given in equation (4.10). The voltage across the resistor leads V_a by the angle ϕ of eqn. (4.8). At high frequencies, there is little voltage across the capacitor because its reactance is $1/\omega C$; so most of the voltage is across the resistor and the angle ϕ is small. The voltage across the resistor leads the applied voltage because this current is charging the capacitor.

The voltage across the capacitor is given by eqn. (4.11) and lags the applied voltage by the angle θ of eqn. (4.9). This is because V_C is proportional to charge and has to wait for the capacitor to charge. A small capacitor charges quickly, so it presents only a small impedence. At high frequencies, $\theta \to 90°$. Conversely at low frequency the capacitor blocks current and most of the voltage is across the capacitor. Then $\theta \to 0$.

Bode plots

Figure 4.8 shows the dependence of $|V_C|$, ϕ and $|V_R|$ on ω. It is convenient to use log scales for both $|V|$ and ω, so as to compress a large range of values on to a single graph. This diagram is called a Bode plot. At low frequency, $|V_C| \simeq V_0$ and the low pass filter transmits the input voltage to the output across C. At low frequencies $\theta \to 0$ and the output is in phase with the input; this is because

Copyright © 2005 IOP Publishing Ltd.

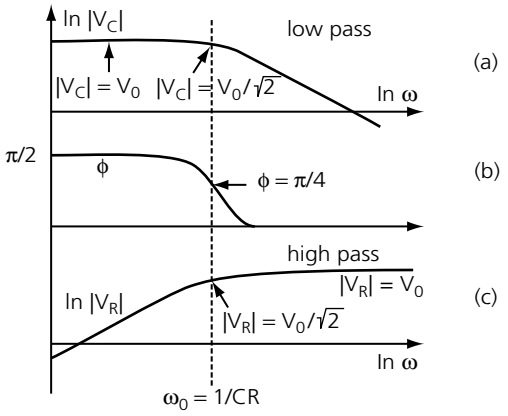


Fig. 4.8. Bode plots for the CR filter.

$V_R \ll V_C$ and $V_C \simeq V_a$ at low frequencies. Conversely, the high pass filter transmits the signal at high frequency, from equation (4.10); at high frequencies $\phi \to 0$, so the output of the high pass filter, V_R, is again in phase with the applied voltage V_a. Here $V_C \ll V_R$ and $V_R \simeq V_a$.

Away from the top of either curve (a) or (c), a phase difference develops between input and output of the filter. As ω rises from 0, $|V_C|$ falls to $V_0/\sqrt{2}$ when $\omega = \omega_0 = 1/CR$; also $\phi \to \pi/4$ (45°) at this frequency. It is no accident that this value of ω is equal to $1/\tau$, where $\tau = CR$ is the time constant of the CR transient (equation 3.9). Later, Fourier's theorem will reveal an intimate relation between the response of any circuit to AC and to a step function.

It is well worth measuring the Bode plot for yourself on the oscilloscope, using say $R = 1\text{k}\Omega$ and $C = 0.1\mu\text{F}$. Vary ω and observe $|V_C|$ and $|V_R|$. The phase difference ϕ between the applied voltage and V_R may be observed by displaying V_{AB} and V_{CD} of figure 4.6(b) on the two traces of an oscilloscope.

Decibels

The vertical scale of the Bode plot is usually expressed in **decibels** (a notation derived historically from loudspeakers). This notation may at first sight look complicated, but is nonetheless useful. The ratio of signal power at the input and at the output of the filter circuit is V_0^2/V_{CD}^2. The quantity

$$r = \log_{10}(V_0^2/V_{CD}^2)$$

Table 4.1 Conversion of attenuations to decibels.

V_0/V_{CD}	1	$\sqrt{2}$	2	3	4	8	10	100	1000
Attenuation(db)	0	3	6	9.5	12	18	20	40	60

Copyright © 2005 IOP Publishing Ltd.

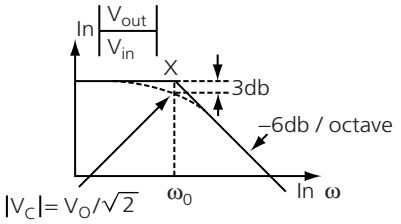


Fig. 4.9. An approximation to the Bode plot of a CR filter circuit.

is expressed in bels (after Alexander Graham Bell); alternatively in decibels (db),

$$r = 10\log_{10}(V_0^2/V_{CD}^2)\ \text{db} = 20\log_{10}(V_0/V_{CD})\ \text{db}.$$

A factor 2 in V_0/V_{CD} corresponds to $r = 20\log_{10}2 = 20 \times 0.3010 \simeq 6$ db. Table 4.1 illustrates the conversion from V_0/V_{CD} to db. Every factor 2 in V_0/V_{CD} corresponds to 6 db and every factor 10 in V_0/V_{CD} to 20 db.

Suppose a signal is multiplied by factors f_i in several successive pieces of electronics and $V_{out} = f_1 f_2 f_3 \ldots V_{in}$. The attenuations or amplifications *add* if expressed in decibels:

$$r = 20\log_{10}(V_{out}/V_{in}) = 20\log_{10}(f_1 f_2 f_3 \ldots)\ \text{db}$$

$$= r_1 + r_2 + r_3 + \ldots \text{where} \qquad r_i = 20\log_{10} f_i\ \text{db}.$$

For example, if there are attenuations by factors 2, 2 and 4 in successive circuits, the net attenuation is (6 + 6 + 12) db = 24 db. This is a valuable property of Bode plots. This addition rule is also true for the phase angles ϕ as demonstrated later using complex numbers.

The horizontal scale of figure 4.8 is expressed in **decades** or **octaves**. A factor 10 in frequency is one decade and $\log_{10}\omega$ increases by 1; a factor 2 in frequency is one octave and $\log_{10}\omega$ increases by 0.3. For frequencies well above ω_0, equation (4.11) reads

$$V_C \simeq V_0/\omega CR$$

$$\log_{10}(V_C/V_0) \simeq -\log_{10}(CR) - \log_{10}\omega.$$

This is a straight line relation on the Bode plot, figure 4.9. The slope is −6 db/octave or −20 db/decade. It is often adequate to approximate the behaviour of the filter by the two straight lines of figure 4.9 meeting at the point X at frequency ω_0. This is called the **corner frequency**. At this frequency the true curve is actually a factor $1/\sqrt{2}$ or 3 db below X. Exercise 10 shows that two cascaded filters give a slope at high frequencies of −12 db/octave or −40 db/decade. For n filters in sequence, the slope is $-6n$ db/octave or −20n db/decade.

Copyright © 2005 IOP Publishing Ltd.

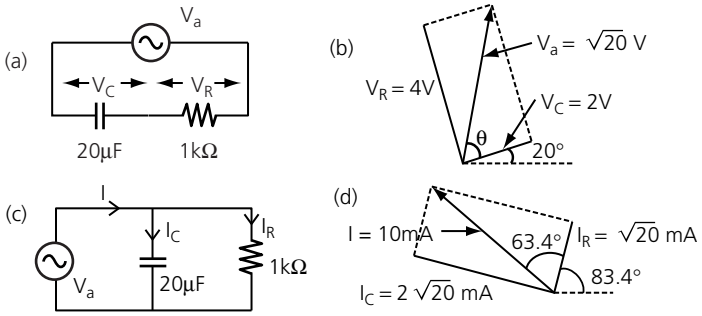


Fig. 4.10. Worked examples on phasors.

Worked examples on phasors

1) Suppose V_C across the 20 μF capacitor in figure 4.10(a) is $2\cos(100t+20°)$ V. What is the applied voltage V_a? To find V_R, the current must be determined. This leads V_C by 90°; its magnitude $\omega C V_C = 2\times100\times2\times10^{-5} = 4\times10^{-3}$ A; it has phase angle 110°. The voltage across the 1 kΩ resistor is 4 V at the same angle. The phasor diagram is shown in figure 4.10(b). The applied voltage V_a is the vector sum of V_C and V_R; it has magnitude $\sqrt{20}$ V and leads V_C by angle $\theta = \tan^{-1} 2 = 63.4°$. So V_a is $\sqrt{20}\cos(100t + 83.4°)$ V.

2) Suppose the same voltage V_a is applied to the same components in parallel, figure 4.10(c). What total current flows? Firstly, $I_R = V_a/R = \sqrt{20}\angle 83.4°$ mA. Next, the magnitude of I_C is $\omega C V_a = 2\sqrt{20}$ mA and it leads the voltage by 90°. The phasor diagram for these currents is shown on figure 4.10(d). The total current is $10\cos(100t + 146.8°)$ mA.

In Chapter 6, an algebra will be developed which avoids drawing these diagrams; nonetheless, they are useful for understanding simple circuits, and you are recommended to try examples from the exercises at the end of the chapter.

4.6 Power Factor

The instantaneous power P dissipated in the CR circuit is

$$P = V_a I = V_0\cos(\omega t - \phi) I_0 \cos\omega t$$
$$= \tfrac{1}{2} V_0 I_0 \{\cos(2\omega t - \phi) + \cos\phi\}.$$

Averaged over a complete cycle, this becomes

$$\boxed{\bar{P} = \tfrac{1}{2} V_0 I_0 \cos\phi = V_{RMS} I_{RMS} \cos\phi.} \tag{4.12}$$

The factor $\cos\phi$ is called the **power factor**.

Copyright © 2005 IOP Publishing Ltd.

An important result is that $\cos\phi = 0$ for a capacitor alone; then $\bar{P} = 0$. Energy is held in the capacitor twice per cycle as it charges first one way then the other, but no power is dissipated. Another way of expressing $\bar{P}$ is, using $\cos\phi = V_R/V_0$, figure 4.7(a),

$$\bar{P} = \tfrac{1}{2} V_R I_0 = \tfrac{1}{2} V_R I_R. \qquad (4.12a)$$

This demonstrates that power is dissipated purely by the resistor.

4.7 Amplifiers

Most amplifiers have an input which behaves simply as a resistance at low frequencies. Ideally, they can be DC coupled to a source. However, in practice internal operation of the amplifier may involve DC voltages which would upset the source or vice versa, and it is quite common to insert a large decoupling capacitor C_1 between source and amplifier, as in figure 4.11(a). The DC levels in the source and the amplifier may then be different. The signal reaching the amplifier is cut off at low frequencies by C_1, making a high pass filter as in figure 4.6(b). It drops by a factor $\sqrt{2}$ at a frequency $\omega_1 = 1/C_1(R_1 + R_2)$. To get a feeling for numbers, suppose the source has an output resistance $R_1 = 2$ kΩ and the amplifier has input resistance $R_2 = 1$ kΩ. These are typical of a simple transistor amplifier. If the circuit is to amplify audio signals down to say $\omega_1 = 100$ rad/s, we require $C \geq 1/(R_1 + R_2)\omega_1$ or 3.3 μF. For frequencies well above ω_1, the reactance of C_1 can be neglected completely and the circuit behaves as if C_1 is absent.

Now consider high frequencies. The amplifier always has some input capacitance C_2, and eventually at high frequencies the reactance of this capacitance will become small compared to R_2 and the input to the amplifier will drop. The high frequency behaviour of the circuit will be analysed, first using Thevenin's theorem, secondly algebraically. In both cases C_1 is neglected.

Suppose the circuit is broken at points CD and the circuit across the capacitor C_2 is replaced by its Thevenin equivalent. When CD is open circuit, $V_{EQ} = V_S R_2/(R_1 + R_2)$. The resistance of the equivalent circuit is obtained by shorting V_S, with the result R_1 in parallel with R_2; so $R_{EQ} = R_1 R_2/(R_1 + R_2)$. The

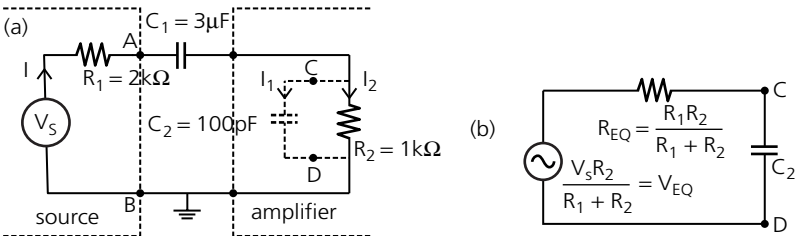


Fig. 4.11. (a) An amplifier AC coupled to a source, (b) its high-frequency equivalent circuit.

Copyright © 2005 IOP Publishing Ltd.

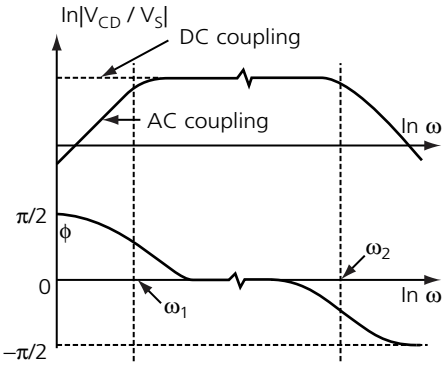


Fig. 4.12. Bode plot for the amplifier.

Thevenin equivalent, figure 4.11(b), is a low pass filter, and the overall behaviour of the amplifier is shown in figure 4.12. At high frequencies, the voltage at CD falls by $1/\sqrt{2}$ when $\omega_2 = 1/C_2R_{EQ}$. Again, to get a rough idea of numbers, suppose $C_2 = 100$ pF, a value typical of an ordinary transistor; and $R_{EQ} = \frac{2}{3}$ kΩ. Then $\omega_2 = 15 \times 10^6$ rad/s.

As a further worked example on phasor diagrams, the high frequency behaviour of the amplifier will be repeated algebraically. This example will illustrate how to treat the parallel combination of C_2 and R_2 using phasors; it is a little awkward and in Chapter 6 a superior algebraic method will be introduced, so do not labour too long over the present example.

Let the output voltage V_{CD} across this parallel combination be $V_2 \cos \omega t = I_2R_2$, and let the source voltage be $V_S \cos(\omega t + \theta)$. Then from figure 4.11(a),

$$V_S \cos(\omega t + \theta) = R_1 I + R_2 I_2 = R_1 I_1 + (R_1 + R_2) I_2$$

$$V_{CD} = I_2 R_2 = (1/C_2) \int I_1 \, \mathrm{d}t.$$

The current I_1 leads I_2 by 90^o, so the vector diagram for V_S is as shown in figure 4.13. The horizontal component is due to I_2 and the vertical component is due to I_1. The output voltage V_{CD} is in phase with I_2 and lags V_S by angle θ where

$$\tan\theta = \frac{R_1\omega C_2}{(R_1 + R_2)/R_2} = \frac{R_1 R_2}{R_1 + R_2}\omega C_2$$

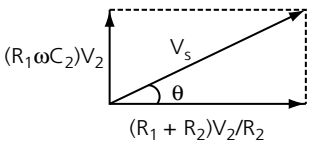


Fig. 4.13. Phasor diagram for V_S and V_2.

Copyright © 2005 IOP Publishing Ltd.

and the magnitudes of V_S and V_2 are related by

$$|V_S|^2/|V_2|^2 = (\omega R_1 C_2)^2 + \{(R_1 + R_2)/R_2\}^2 \tag{4.13}$$

$$\text{or} \qquad |V_2| = \{R_2/(R_1 + R_2)\}|V_S|/\left[1 + \left(\frac{R_1 R_2}{R_1 + R_2}\omega C_2\right)^2\right]^{1/2}.$$

These results are similar to equations (4.9) and (4.11), except that R has been replaced by $R_{EQ} = R_1 R_2/(R_1 + R_2)$ and the voltage at frequencies below ω_2 is $V_2 = V_{EQ} = R_2 V_S/(R_1 + R_2)$. This gives an algebraic *derivation* of the Thevenin equivalent circuit of figure 4.11(b). In this example, the simpler derivation comes from *assuming* Thevenin's theorem, as in figure 4.11(b).

4.8 Bandwidth

The gain of the amplifier on the flat part of figure 4.12 at intermediate frequencies is called the **mid-band gain**. The frequencies ω_1 and ω_2 at which the output drops to $1\sqrt{2}$ of its maximum are called the **lower** and **upper cut-off frequencies**. The range of ω from ω_1 to ω_2 over which the response curves are approximately flat is known as the **bandwidth** of the circuit. This is a very important quantity. An oscilloscope, for example, has a large bandwidth, so that it displays signals over a wide frequency range. Generally speaking, it is hard to achieve large bandwidth, but easy for the user to restrict it for particular applications (e.g. selecting a radio station). Manufacturers try to produce circuits and transistors with large bandwidth for multipurpose use.

With the numerical values given for the amplifier of figure 4.11, $\omega_1 = 10^2$ and $\omega_2 = 1.5 \times 10^7$ rad/s. Suppose an audio amplifier is to be optimised instead for the range $\omega = 10^2$ to 10^5 rad/s. The bandwidth may be moved down to centre on this range by using values of R_1 and R_2 a factor 10 or so larger. On the other hand, suppose an amplifier is required for TV frequencies (100 MHz) or the even higher frequencies used in computers or radio astronomy. To achieve a bandwidth reaching such frequencies, it is essential to keep C_2 small and also R_1 and R_2. Development of transistors is aimed at reducing capacitance to achieve higher speeds, and gallium arsenide devices now work up to 10^{11} Hz.

4.9 Noise and Bandwidth

No electrical circuit is ever completely quiescent. Even if there is no voltage applied to it, thermal fluctuations give rise to fluctuating voltages. This is called **Johnson noise** or **thermal noise**. If you turn the sensitivity of an oscilloscope up

Copyright © 2005 IOP Publishing Ltd.

to the maximum, you will usually see such noise on the trace. In a resistance R, it may be shown that these fluctuations give rise to a mean power

$$V_{RMS}^2/R = 4kT\ \Delta f \tag{4.14}$$

in a frequency interval Δf. Here, k is Boltzmann's constant and at room temperature $kT \simeq 1.4 \times 10^{-23} \times 300$ joules. If $R = 10^5\ \Omega$ and $\Delta f = 10^6$ Hz for example, $V_{RMS} = 40\ \mu$V. The power is uniformly spread over all frequencies, and this type of noise is therefore called **white noise**. In detecting weak signals, it is necessary to restrict the bandwidth to a small range, to reduce noise.

A second type of noise is **shot noise**. This arises from the fact that current is not continuous, but is carried by electrons whose charge is quantised in units of $e = 1.6 \times 10^{-19}$ C. Imagine, for example, an electron crossing the gap of a capacitor. As it approaches one plate, it repels electrons in the plate and induces a surge of current. If mean current I flows, the *RMS* fluctuation in that current is

$$I_{RMS}^2 = 2eI\ \Delta f \tag{4.15}$$

and again the effect is uniformly distributed over all frequencies. If $I = 1$ mA and $\Delta f = 10^6$ Hz, $I_{RMS} = 1.8 \times 10^{-8}$ A.

At what current is $I_{RMS} = I$ if $\Delta f = 10^6$ Hz? To what does this correspond in terms of electrons/unit bandwidth?

These two sources of noise are inescapable, though they can be minimised by suitable choice of circuit parameters (R, I and temperature T). In addition, there may be stray pick-up from the mains, corona discharge from power lines, car ignition, earth loops and so on. These can be minimised by shielding. There is also a further source of noise called **contact** or **1/f noise**. It originates from poor contacts within conductors; it is much worse in a carbon-composition resistor, for example, than in a metal-film resistor. Empirically, it is given by

$$I_{RMS}^2 = K I\ \Delta f/f \tag{4.16}$$

where K depends on the material. Because its frequency spectrum is proportional to $1/f$, it dominates at low frequency, typically below 1 kHz in practice.

How do noise signals of different frequencies and from different sources combine? Now

$$V_{total} = V_1 + V_2 + V_3 + \ldots$$

$$V_{total}^2 = V_1^2 + V_2^2 + V_3^2 + \ldots + 2V_1V_2 + 2V_1V_3 + 2V_2V_3 + \ldots.$$

Because the signals are random, there is no correlation between V_1 and V_2 and so on. The average value of V_1V_2 is zero and likewise for other cross-terms. So values of V_i^2 add, i.e. noise powers add.

Copyright © 2005 IOP Publishing Ltd.

4.10 Summary

A capacitor has reactance $1/\omega C$; it blocks DC and transmits high frequency signals. AC current through a capacitor leads the voltage across it by 90°. The voltage across the capacitor is proportional to charge and therefore lags the applied voltage. The lower cut-off frequency of a CR filter is $\omega_1 = 1/CR_s$, where R_s is the *series* resistance; the upper cut-off frequency of a filter or amplifier is $\omega_2 = 1/CR_p$, where R_p is the resistance in *parallel* with the capacitor. The bandwidth of the circuit, from ω_1 to ω_2 is the range of ω over which its output is roughly constant and within a factor $1/\sqrt{2}$ of its maximum value. The bandwidth of an amplifier is in practice very important. The phase of the applied voltage is 45° ahead of the output of the amplifier at ω_2 and 45° behind it at ω_1 (figure 4.12). The power dissipated in a CR filter circuit is $V_{RMS}I_{RMS}\cos\phi$ and is dissipated purely by the resistor; capacitors store energy temporarily, but do not dissipate it. The magnitudes of thermal noise and shot noise are proportional to bandwidth and are uniformly distributed with ω; noise powers add incoherently.

4.11 Exercises

1. For a 50 Hz signal, what is ω? What is the frequency of the light from a bulb fed by the mains? Why is it difficult to generate very high frequency signals?

2. What is the impedance of a 1μF capacitance at (a) 50 Hz, (b) 10^6 Hz? Design a high-pass filter with a cut-off at $\omega = 10^6$ rad s^{-1}, using a 50 ω resistor. Sketch the Bode plot for this filter.

   For subsequent problems, begin by drawing a phasor diagram and marking the angles between applied voltage (or current) and those across resistors and capacitors. For problems involving more than one resistor or capacitor, Thevenin's theorem is often useful in breaking the problem up into steps. If you find yourself struggling with the more difficult problems, Chapter 6 will provide a more transparent way of writing the algebra. If you are already familiar with complex numbers and complex exponentials, try using them on these problems. If necessary, come back to them later after studying Chapter 6.

3. A resistance $R = 120\ \Omega$ and a capacitance $C = 12.5\ \mu$F are connected in series; current $I = 1.5\cos(5000t + 25°)$ mA flows through them. What is the voltage across each component and the total applied voltage? What is the power factor of the circuit and what power is dissipated in each component? (Ans: $V_R = 180\cos(5000t + 25°)$ mV, $V_C = 24\cos(5000t - 65°)$ mV, $V_a = 181.6\cos(5000t + 17.4°)$ mV; 0.9912; $\bar{P}_R = 135\ \mu$W, $\bar{P}_C = 0$.)

Copyright © 2005 IOP Publishing Ltd.

4. A resistor $R = 1\ \text{k}\Omega$ is in series with an unknown capacitor C. The voltage V_R across the resistor is $12\cos 10^4 t$ V, and V_R leads the applied voltage by 33.7°. What is C? (Ans: 0.15 μF.)

5. What is the current in each of the three circuits of figures 4.14(a)–(c) and what are the voltages across each of the 6 components (after transients have died away), supposing that the applied voltage is $12\cos\omega t$ V in each case, with $\omega = 10^3$ rad/s? (Ans: (a) $150\cos\omega t\ \mu$A, $4.95\cos\omega t$ V, $7.05\cos\omega t$ V; (b) $-9\sin\omega t$ mA, $9\cos\omega t$ V, $3\cos\omega t$ V; (c) $7.68\cos(\omega t + 39.8°)$ mA, $7.68\cos(\omega t - 50.2°)$ V, $9.22\cos(\omega t + 39.8°)$ V.)

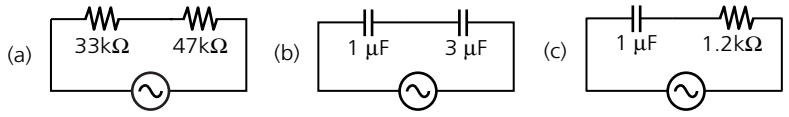


Fig. 4.14.

6. If in each of figures 4.14(a)–(c) the components are rearranged in parallel, what is the total current in each circuit and the current through each component? (Ans: (a) $619\cos\omega t\ \mu$A, $364\cos\omega t\ \mu$A, $255\cos\omega t\ \mu$A, (b) $-48\sin\omega t$ mA, $-12\sin\omega t$ mA, $-36\sin\omega t$ mA; (c) $15.6\cos(\omega t + 50.2°)$ mA, $-12\sin\omega t$ mA, $10\cos\omega t$ mA.)

7. The AC input of an oscilloscope has a capacitor C in series with the input resistance R of the instrument. If $R = 10^6\Omega$ and DC input is to fall with a time constant of 0.1 s, what magnitude is required for R? What is then the cut-off angular frequency ω and the corresponding frequency f? (Ans: 0.1μF, $\omega = 10$ rad s^{-1}, $f = 10/2\pi\ s^{-1}$.)

8. The voltage source in figure 4.15 is $V\cos\omega t$. Find the power dissipation in the circuit and show that it is the same for two different values of R. What is the value of R for maximum power dissipation? The magnitude of R is adjusted to this value and an equal resistance is placed in parallel with C. Find the Thevenin equivalent across C and hence the current flowing in C. (Ans: $\bar{P} = \frac{1}{2}V^2R/\{R^2 + (1/\omega C)^2\}$; $1/\omega C$; $(\omega CV/\sqrt{5})\cos(\omega t + 63.4°)$.)

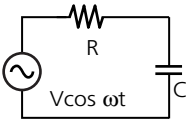


Fig. 4.15.

9. (Westfield). Explain the meaning of the decibel scale and show how it can be expressed in terms of voltage ratios. A switched attenuator has

Copyright © 2005 IOP Publishing Ltd.

two controls calibrated in decibels, in tens and units respectively. To what values must these controls be set if a reduction in voltage by a factor of twenty is required? (Ans: 26 db.)

10. Find the output voltage from the two-stage filter in figure 4.16 for frequencies large compared with $1/C_1R_1$ and $1/C_2R_2$. Hint: find the current I through R_1 ignoring the effect of R_2 and C_2; for this current find V_X, then find I_2 hence an approximate value of the output voltage. (Ans: $V_{out}/V_S \simeq -1/\omega^2 C_1 C_2 R_1 R_2$.)

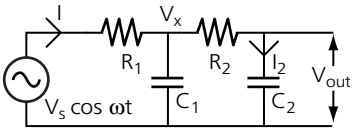


Fig. 4.16.

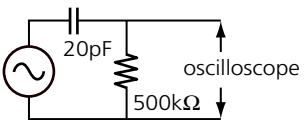


Fig. 4.17.

11*. A voltage $20\sin\omega t$ V is applied to the circuit in figure 4.17 with $\omega = 10^5$ rad/s; the voltage across the 500 kΩ resistor is measured using an oscilloscope with input resistance R_{in} = 1 MΩ and parallel input capacitance C_{in} = 1 pF. (a) Find the voltage across the resistor, ignoring the loading by the oscilloscope. (b) Find the voltage across this resistor including R_{in} but ignoring C_{in}. (c) Use a Thevenin equivalent for the circuit feeding C_{in}. What is then the observed voltage? By how much is the observed phase angle wrong due to the loading by the oscilloscope? (Ans: (a) $14.1\sin(\omega t + 45°)$, (b) $11.1\sin(\omega t + 56.3°)$, (c) $10.9\sin(\omega t + 55.0°)$ V; $+10.0°$.)

Copyright © 2005 IOP Publishing Ltd.

5

Inductance

5.1 Faraday's law

Suppose a coil is wound around a cylindrical support or 'former' as in figure 5.1. This arrangement is called a **solenoid**. If a current is passed through the coil, a magnetic field **H** is generated along the length of the solenoid. If the solenoid has 100 turns/cm, a suitable current is $I = 20$ mA. The fundamental experiments of Ampère, who established the laws governing the magnetic field, are described in textbooks of electromagnetism. The points of concern here are that H is parallel to the axis of the solenoid and is proportional to current and to the number of turns m:

$$H \propto mI. \tag{5.1}$$

Mutual inductance

Next suppose two separate coils are brought close together, as in figure 5.2. A current I_1 is fed through coil 1 and the second is connected to an oscilloscope. When I_1 is steady no signal is observed on the oscilloscope. But if I_1 changes (e.g. by breaking the first circuit), a signal is observed from the second coil. This observation was first made by Faraday, using a sensitive galvanometer. He studied the dependence on I_1, on the orientation of the two coils and on movement of the second coil with respect to the first. He concluded that a *voltage* (rather than a current) is generated with a magnitude given by the rate of change of magnetic flux F in the second coil. This flux is defined as $F = H_n A$, where A is the area of the coil and H_n is the component of magnetic field normal to the plane of the coil; more exactly, if H_n varies over the area, $F = \int H_n \, \mathrm{d}A$. Faraday's law of electromagnetic induction is one of the five basic laws of electromagnetism and states that the induced voltage V is

$$\boxed{V = -\frac{\mathrm{d}F}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\int H_n \, \mathrm{d}A.} \tag{5.2}$$

Copyright © 2005 IOP Publishing Ltd.

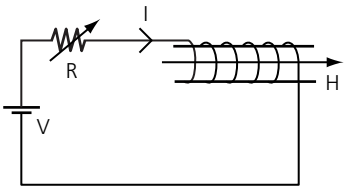


Fig. 5.1. The field of a solenoid.

The minus sign indicates that the induced voltage has a sign such as to produce a current which opposes the change of field.

From equation (5.1), $H_1 \propto I_1$ and thus the voltage in coil 2, $V_2 \propto \mathrm{d}I_1/\mathrm{d}t$. The proportionality constant M is called the **mutual inductance**:

$$\boxed{V_2 = M\mathrm{d}I_1/\mathrm{d}t.} \tag{5.3}$$

It is measured in **henries** (H). With M in these units,

$$V\,(\text{volts}) = M\,(\text{henries})\frac{\mathrm{d}I\,(\text{amps})}{\mathrm{d}t\,(\text{seconds})}. \tag{5.4}$$

Experiment and theory both demonstrate that the behaviour of the two coils is reciprocal: if a current I_2 varies in the second coil, a voltage V_1 is induced in the first with

$$\boxed{V_1 = M\mathrm{d}I_2/\mathrm{d}t} \tag{5.5}$$

and precisely the same value of M.

Magnetic fields are greatly enhanced in some ferromagnetic materials: iron, cobalt and nickel. If a cylindrical iron core is inserted down the axes of the two coils of figure 5.2, the observed voltages are enhanced by a large factor.

5.2 Self-inductance

The magnetic field produced by one coil can induce a voltage in a second. Why not in itself too? If you join coil 1 and coil 2 of figure 5.2 at points X and Y, it should not make any difference.

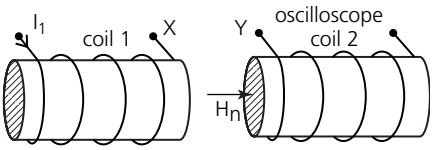


Fig. 5.2. Mutual inductance.

Copyright © 2005 IOP Publishing Ltd.

Observations described in the next section demonstrate that a changing current I in a coil indeed generates a voltage V across this same coil where

$$\boxed{V = L\mathrm{d}I/\mathrm{d}t,} \tag{5.6}$$

and L is called **self-inductance**. The sign of the voltage is such as to produce a current resisting the change. (If it had the opposite sign, there would be a runaway situation where any change of current in a coil would produce a continuously increasing current).

The remainder of this chapter explores the response of a coil or **inductor** first to a voltage step then to an AC voltage. There is a close parallel to the behaviour of a capacitor. Mutual inductance plays little role in most electronic circuits but is important in power supplies. The discussion of these is postponed to chapter 16.

5.3 LR Transient

An inductance resists a change of current. If a voltage is applied across an inductor by closing the switch in figure 5.3, the current will grow only gradually and a transient is observed as current climbs to the value V/R in the absence of the inductor. This transient turns out to be very similar to the *CR* transient for charging a capacitor. In the present case, the voltage V_L across the inductor falls exponentially with time:

$$V_L = V_0 \mathrm{e}^{-t/\tau} \qquad \text{where} \qquad \tau = L/R$$

This is sketched on figure 5.3

When the switch is closed at $t = 0$, the initial current is zero, and the full applied voltage V appears across the inductor. As the current grows, V_L falls as V_R grows: $V_L = V - IR$. Setting $V_L = L\mathrm{d}I/\mathrm{d}t$,

$$V = L\mathrm{d}I/\mathrm{d}t + IR. \tag{5.7}$$

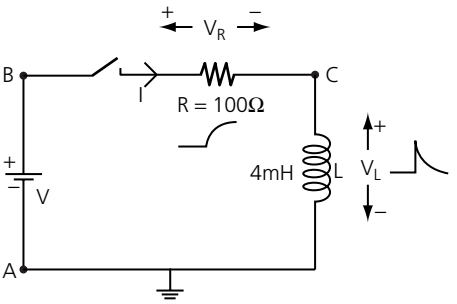


Fig. 5.3. Series LR circuit.

Copyright © 2005 IOP Publishing Ltd.

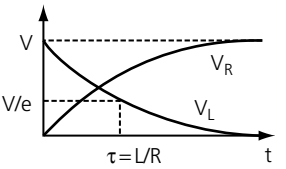


Fig. 5.4. V_R and V_L v. t.

This is a first-order differential equation like the one in Chapter 3, except for the appearance of the constant V on the left-hand side. It may be solved by the trial substitution

$$I = A\exp(-t/\tau) + B;$$

A and B are constants. The constant term B is required to give Ohm's law $I = V/R$ as $t \to \infty$. Differentiating I with respect to t,

$$\mathrm{d}I/\mathrm{d}t = -(A/\tau)\exp(-t/\tau);$$

substituting in (5.7)

$$V = (-LA/\tau)\exp(-t/\tau) + AR\exp(-t/\tau) + BR.$$

This equation is satisfied for any A and any t if

$$\boxed{\tau = L/R} \tag{5.8}$$

and $B = V/R$. With L in henries and R in ohms, τ is in seconds.

To find the constant A, the initial conditions must be used. The condition $I = 0$ at $t = 0$ requires $A = -B$, so finally

$$\boxed{I = (V/R)\{1 - \exp(-t/\tau)\}.} \tag{5.9}$$

Figure 5.4 shows the voltage against time across the resistor and across the inductor. A large value of L hinders the creation of the current; for this reason an inductor is often called a **choke**.

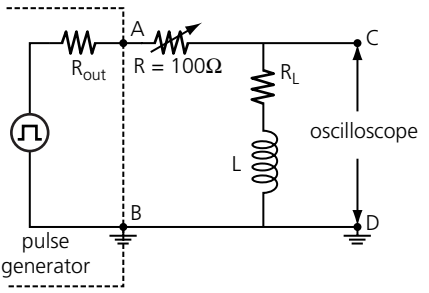


Fig. 5.5. Circuit to observe V_L.

Copyright © 2005 IOP Publishing Ltd.

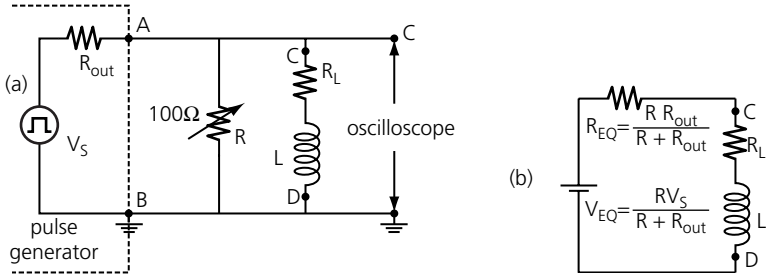


Fig. 5.6. (a) Variant if R_{out} is large, (b) its equivalent circuit.

If the initial current is I_0 instead of zero, how are equation (5.9) and figure 5.4 affected?

From electromagnetic theory, the self inductance of a long solenoid of volume V and n turns/metre is $L = 4\pi \times 10^{-7} n^2 V$ Henries, so you may observe these phenomena using a coil of about 1000 turns wound on a former of 2 cm diameter and 10 cm long. This produces a self-inductance of about 4 mH $= 4 \times 10^{-3}$ H. The battery of figure 5.3 may be replaced by a square-wave generator, as in figure 5.5. The coil will have a resistance R_L of a few ohms, which can be measured with a multimeter. The voltage V_L is observed with an oscilloscope connected across CD. It is necessary to allow for the voltage drop IR_L if this is significant. It is instructive to insert a cylinder of soft iron inside the former of the inductor and observe the change in L.

In interpreting the observations, it is necessary to remember that R of equation (5.8) is replaced by the total series resistance $(R + R_{out} + R_L)$ of figure 5.5. Providing the output resistance of your pulse generator is small ($< 100\Omega$), τ will be $\geq 4 \times 10^{-5}$ s and the current rise can be observed readily. Suppose, however, your generator has an output impedance which is large ($> 1000\Omega$). In this case, τ is inconveniently small. The remedy lies in figure 5.6(a), where the variable resistor R is put in parallel with the coil. The Thevenin equivalent circuit across CD, figure 5.6(b), has R_{EQ} given by the parallel combination R and R_{out}, and can be made arbitrarily small by reducing R. The time constant becomes $\tau = L/(R_L + R_{EQ})$. The penalty is that the voltage of the Thevenin equivalent circuit is $V_{EQ} = RV_S/(R + R_{out})$ and this decreases for small R. However, by turning up the pulse generator to give an output of say 10 V, you can still observe a reasonable size of signal across CD.

5.4 Energy Stored in an Inductor

If a current I is established in an inductor, the work done is given by

$$E = \int V I \, \mathrm{d}t = \int L \frac{\mathrm{d}I}{\mathrm{d}t} I \, \mathrm{d}t = \int L I \, \mathrm{d}I$$

Copyright © 2005 IOP Publishing Ltd.

or

$$\boxed{E = \tfrac{1}{2}LI^2.} \tag{5.10}$$

For large currents in a transformer coil, this stored energy can be quite large. If a sudden break occurs in the circuit, the induced voltage can be enough to cause a spark. Ignition coils of cars work on this principle. In circuits where you do NOT wish this to happen, you should put a parallel resistor across the inductor to provide a discharge path (see exercise 11).

Can you understand in terms of equation (5.10) why a small time constant τ goes with large R?

5.5 Stray Inductance

Inductive effects are not restricted to the geometry of a coil. They arise for a conductor of any shape, even a straight wire. Figure 5.7 shows the field around a straight wire. It is at right angles to the current and in the direction of a right-handed screw pointing along the direction of the current. Some of this magnetic field lies within the wire itself. If the current changes, an emf is generated within the wire causing currents which oppose the change. If the current is uniformly distributed in the wire, it can be shown that the self-inductance of length ℓ (metres) is

$$L = \mu_0 \ell / 8\pi$$

where $\mu_0 = 10^{-7}$. Although numerically small, it is not always negligible. At high frequencies, there is also the complication that the current is concentrated near the surface of the wire (the skin effect). Resistors and even wires have such **stray inductance** and leads must be kept as short as possible for high frequencies.

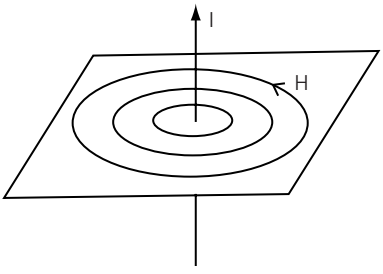


Fig. 5.7. Field of a straight wire.

Copyright © 2005 IOP Publishing Ltd.

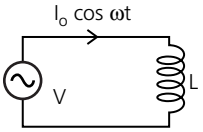


Fig. 5.8. Alternating voltage applied to an inductor.

5.6 Response of an Inductor to Alternating Current

When an alternating voltage is applied to an inductor (figure 5.8),

$$V = L\mathrm{d}I/\mathrm{d}t$$

and if

$$I = I_0 \cos \omega t$$

$$V = -\omega L I_0 \sin \omega t = \omega L I_0 \cos(\omega t + \pi/2).$$

The magnitude of the voltage is $\omega L I_0$ and the reactance of the inductor is ωL. The voltage leads current by $\pi/2$ (figure 5.9). The physical reason is that an inductor hinders current flow, so current follows voltage. Averaged over a cycle, no power is dissipated in an inductor, because of the phase difference of $\pi/2$.

5.7 Phasors

We can extend to inductors the phasor diagrams introduced for capacitors in the previous chapter. Suppose, for example, an AC voltage is applied to a series combination of a resistance R and an inductor (figure 5.10). Let the current in the circuit be $I = I_0 \cos \omega t$. Then, from Kirchhoff's voltage law

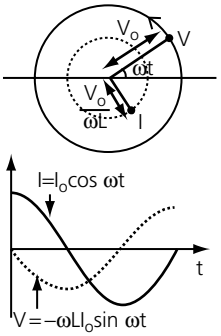


Fig. 5.9. Current in an inductor lags by 90°.

Copyright © 2005 IOP Publishing Ltd.

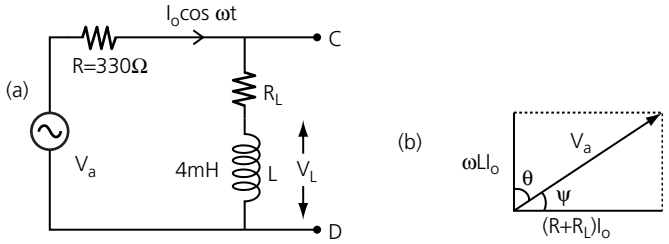


Fig. 5.10. (a) RL series circuit and (b) its phasor diagram.

$$V_a = RI + R_L I + L\mathrm{d}I/\mathrm{d}t$$

$$= (R + R_L)I_0 \cos \omega t + \omega L I_0 \cos(\omega t + \pi/2).$$

The corresponding phasor diagram is shown on figure 5.10(b). The inductor has reactance ωL and gives rise to a voltage $\omega L I_0$ leading the current by 90°. Then

$$V_a^2 = \{(R + R_L)^2 + (\omega L)^2\}^{1/2} I_0^2 \tag{5.11}$$

$$\tan \psi = \omega L/(R + R_L).$$

The applied voltage leads the current by angle ψ, but it lags V_L, the voltage across the inductor, by angle $\theta = (\pi/2 - \psi)$. From equation (5.11), the circuit operates as a high pass filter:

$$|V_L|/|V_a| = \frac{\omega L}{\{(R + R_L)^2 + (\omega L)^2\}^{1/2}} = \frac{1}{\{1 + (1/\omega L)^2 (R + R_L)^2\}^{1/2}}.$$

The cut-off frequency is $\omega_0 = (R + R_L)/L$. It is straightforward to find L by using an oscilloscope to observe V_{CD} and sweeping the frequency of the generator from high to low values to find ω_0; however, the voltage V_{CD} includes IR_L which must be included in the arithmetic if it is significant compared to $\omega L I_0$.

5.8 Summary

Just as it takes time for charge to build up on a capacitor, so it takes time for a current to become established in an inductor: current through an inductor cannot change instantaneously. The time constant is L/R. The reactance is ωL. An inductor transmits DC but hinders AC. The current through an inductor lags the voltage across it by 90°. Averaged over one cycle of AC, no power is dissipated by an inductor. Energy $\frac{1}{2} I^2 L$ is stored in it. If the current is suddenly broken, the voltage $L\mathrm{d}I/\mathrm{d}t$ creates a surge which can damage other components in the circuit (e.g. transistors) unless a low resistance discharge path is provided.

Copyright © 2005 IOP Publishing Ltd.

5.9 Exercises

1. For an inductor and resistor in series, why is the time constant τ small if L is small or R is large? What is the equivalent inductance of figure 5.11? (Ans: 8/3 mH.)

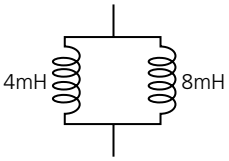


Fig. 5.11.

2. An inductance of 0.1 H carries a current 25 cos $100t$ A. What is the voltage across it? (Ans: $-250 \sin 100t$ V.)

3. A wire wound resistor of 20Ω has a stray inductance of 5×10^{-7} H. Will this appear in series or in parallel with R? At what frequency does this stray inductance become significant and what is then its effect? (Ans: in series, for a sensible result at low frequency; $\omega_0 = 4 \times 10^7$ rad/s; makes a low pass filter and for $\omega > \omega_0$, L dominates.)

4. The voltage across a 4 mH inductor is $12 \sin(1000t + 20°)$ mV; what current flows through it? (Ans: $3 \sin(1000t - 70°)$ mA).

5. A constant voltage of 12 V is applied to a series combination $R = 5\,\Omega$ and $L = 20$ H by closing a switch at time $t = 0$. Find the current v. time and the voltages V_R and V_L across R and L. At what time is $V_R = V_L$? (Ans: $I = 2.4(1 - e^{-t/4})$ A, $V_R = 12(1 - e^{-t/4})$ V, $V_L = 12\mathrm{e}^{-t/4}$ V, 2.77 s.)

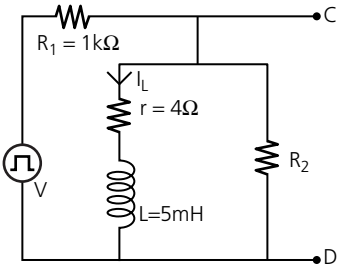


Fig. 5.12.

6**. Show algebraically that in figure 5.12 the current I_L through the inductor is given by

$$\frac{VR_2}{R_1 + R_2} = \frac{R_1 R_2 I_L}{R_1 + R_2} + L\frac{\mathrm{d}I_L}{\mathrm{d}t} + r I_L. \qquad \text{(Thevenin equivalent).}$$

Copyright © 2005 IOP Publishing Ltd.

If the voltage source is a generator of 4 V square pulses, what value of R_2 is required to observe pulses at *CD* with a time constant of 0.1 ms, and what is then the height of the pulses at *CD*? (Ans: $R_2 = 48\ \Omega$, 184 mV.)

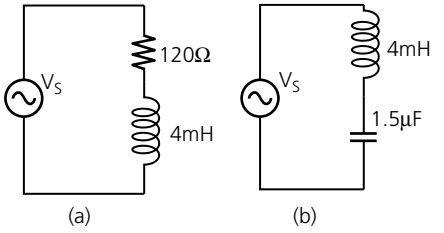


Fig. 5.13.

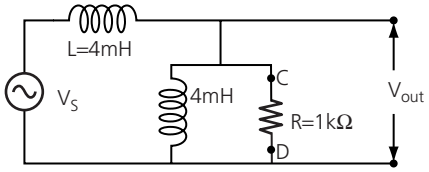


Fig. 5.14.

7. What currents flow in the circuits of figure 5.13(a) and (b) (after transients have died away) if the voltage source is $2\cos(2 \times 10^4 t + 30°)$ V? What are the voltages across individual components? (Ans: (a) $13.9\cos(2 \times 10^4 t - 3.7°)$ mA; $V_R = 1.67\cos(2 \times 10^4 t - 3.7°)$ V; $V_L = 1.11\cos(2 \times 10^4 t + 86.3°)$ V; (b) $42.9\cos(2 \times 10^4 t - 60°)$ mA; $V_L = 3.43\cos(2 \times 10^4 t + 30°)$ V; $V_C = -1.43\cos(2 \times 10^4 t + 30°)$ V.)

8. If the components in each of figures 5.13(a) and (b) are rearranged in parallel, what current flows through each component and what is the total current in each circuit? (Ans: (a) $I_R = 16.7\cos(2 \times 10^4 t + 30°)$ mA, $I_L = 25\cos(2 \times 10^4 t - 60°)$ mA, $30\cos(2 \times 10^4 t - 26.3°)$ mA, (b) $I_L = 25\cos(2 \times 10^4 t - 60°)$ mA, $I_C = 60\cos(2 \times 10^4 t + 120°)$ mA, $I = 35\cos(2 \times 10^4 t + 120°)$ mA.)

9. In the circuit of figure 5.14, what is the output voltage $V_{out}(\omega)$ if $V_S = 10\cos\omega t$ V? Does the circuit behave as a low or high pass filter? What is the cut-off frequency? Hint: consider the Thevenin equivalent across *CD*. (Ans: $V_{out} = \frac{1}{2}V_S R\cos(\omega t - \psi)/\{(\frac{1}{2}\omega L)^2 + R^2\}^{\frac{1}{2}}$ with $\tan\psi = \omega L/2R$; low pass; $\omega_0 = 5 \times 10^5$ rad/s.)

10. In figure 5.15, what is the output voltage $V_{out}(\omega)$? Hint: Let $V_{out} = V_0\cos(\omega t - \psi)$ and find I_L and I_C. (Ans: $V_{out} = (4/2.04)\cos(\omega t - \psi)$ with $\tan\psi = 5$.)

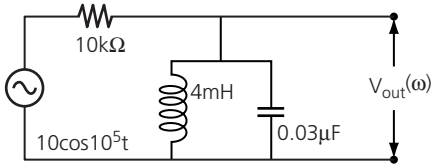


Fig. 5.15.

Copyright © 2005 IOP Publishing Ltd.

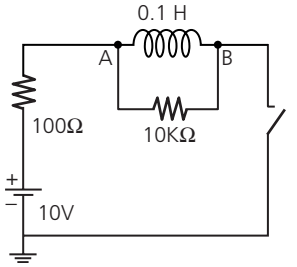


Fig. 5.16.

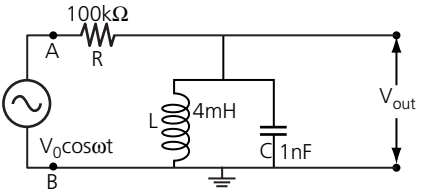


Fig. 5.17.

11**. (QMC). The switch in figure 5.16 has been closed for a long time before it is suddenly opened. Sketch voltage waveforms for points A and B with respect to earth and the waveform of current in the inductor, marking values immediately before and after the opening of the switch and the final values attained after the decay of the transient voltages and currents. What is the time constant of these transients? Assume the inductor is loss free. (Ans: $V_A = V_B = 0$ before; $V_A = 10$ V, $V_B = 1010$ V after; V_B falls exponentially to 10 V with $\tau = 10^{-5}$ s.)

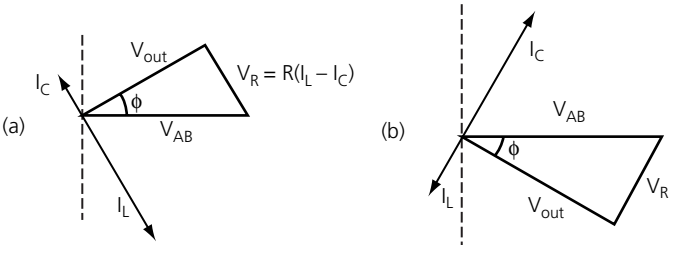


Fig. 5.18.

12. (QMW). ∗ Show that, for the circuit shown in figure 5.17,

$$|V_{out}| = V_0/[1 + \{\frac{CR}{L}(\omega L - 1/\omega C)\}^2]^{1/2}$$

and find the relation between V_{out} and V_0. Find expressions for the currents through the capacitor and inductor and display these currents as phasor diagrams relative to V_{AB} for (a) $\omega < 1/\sqrt{LC}$, (b) $\omega > 1/\sqrt{LC}$, Sketch the Bode plot for V_{out} as a function of frequency, indicating the essential qualitative features. (Ans: (a) and (b) figures 5.18(a) and (b); $V_{out} \to 0$ at low and high frequencies; resonance at $\omega = 1/\sqrt{LC}$ with $V_{out} = V_0$.)

Copyright © 2005 IOP Publishing Ltd.

6

Complex Numbers: Impedance

6.1 Complex Numbers

Complex numbers play a fundamental role in many areas of physics and engineering. They express information about both magnitude and phase in a very simple way. They are central to understanding Fourier transforms (Chapter 15), and it is worth the effort to become fluent in using them. In treating AC circuits, they handle capacitances and inductances in a neat, compact way. The phasor diagrams of the previous two chapters are in fact graphical representations of complex numbers. Usually it is easier to handle circuit problems algebraically than with diagrams, so this chapter will develop and illustrate the algebra. First, we review the pure mathematics; then we apply it to AC circuits.

As a preliminary, consider the properties of 'ordinary' or **real** numbers. Examples are –4, –3, –2 . . . 1, e, π, 10, and so forth. They can be represented by points plotted along a single axis, figure 6.1, called the 'real axis'. The sum or difference or product of any two real numbers is another real number: e.g. $2 + 3 = 5$ and $2 \times 3 = 6$. The square of any real number is a real number: $3^2 = 9$. However, the converse is not always true: $\sqrt{9} = +3$ or –3, but there is no real number whose square is –9.

The notion of the square root of a negative number was discussed in 1575 by Bombelli, but it was Euler who in 1777 introduced a specific notation. Mathematicians use the symbol i for the square root of –1. In electrical engineering, it is conventional to use instead the symbol j, to avoid confusion with currents:

$$\boxed{\mathrm{j}^2 = -1} \tag{6.1}$$

or

$$\boxed{\mathrm{j} = \sqrt{-1}.} \tag{6.2}$$

By this formal device, algebraic expressions may be written for the square roots of negative numbers, e.g. $\sqrt{-9} = 3\mathrm{j}$ or $-3\mathrm{j}$. These are called **imaginary** numbers. Examples are $-4\mathrm{j}$, $-3\mathrm{j}$, $-2\mathrm{j}$, $\ldots\mathrm{j}$, $\pi\mathrm{j}$, and so on. They are distinct from real numbers, but differ only by the appearance of the factor j. They may be plotted

Copyright © 2005 IOP Publishing Ltd.

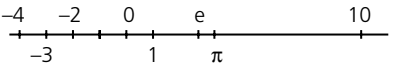


Fig. 6.1. Real numbers.

along a new axis called the 'imaginary axis'. It is conventional to plot the real and imaginary axes at right angles to one another, as in figure 6.2. This plot is called the **Argand diagram** or **complex plane**.

Complex numbers have both real and imaginary parts. An example is the point $Z = 4 + 5\mathrm{j}$ shown on figure 6.2. Real numbers, e.g. (4,0) are a special case of complex numbers, as are imaginary numbers, e.g. (0,5) $\equiv 5\mathrm{j}$.

Complex numbers obey the same rules of addition, subtraction, multiplication and division as real numbers. For example:

$$\text{addition} : (2 + 3\mathrm{j}) + (5 + 9\mathrm{j}) = 7 + 12\mathrm{j}$$

$$\text{subtraction} : (7 + 12\mathrm{j}) - (5 + 9\mathrm{j}) = 2 + 3\mathrm{j}$$

$$\begin{aligned}\text{multiplication} : (A + \mathrm{j}B)(C + \mathrm{j}D) &= (AC + \mathrm{j}^2 BD) + \mathrm{j}(BC + AD) \\ &= (AC - BD) + \mathrm{j}(BC + AD). \qquad (6.3)\end{aligned}$$

(Division will be dealt with below.) These examples demonstrate that the sums and products of complex numbers are themselves complex numbers. They are the simplest possible generalisation of the notion of a real number. The only fresh ingredient is the definition $\mathrm{j}^2 = -1$.

Modulus and phase

The point Z of figure 6.2 may alternatively be described in terms of R and θ (figure 6.3). The distance R of the point from the origin is called the **modulus** of Z and is written $|Z|$. The angle θ is called its **phase** or sometimes the **argument** arg of Z. By convention, θ is measured from the real axis and is positive towards

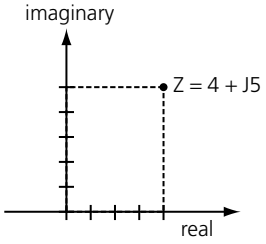


Fig. 6.2. A complex number.

Copyright © 2005 IOP Publishing Ltd.

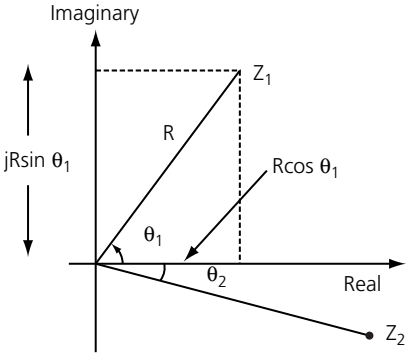


Fig. 6.3. The R, Θ representation.

the imaginary axis; thus θ_1 is positive for Z_1 in figure 6.3 and θ_2 is negative for the point Z_2.

In terms of these quantities,

$$Z = R(\cos\theta + \mathrm{j}\sin\theta). \tag{6.4}$$

This is a very useful form for expressing complex numbers. We say that 'the real part of Z is $R\cos\theta$' or

$$Re\ Z = R\cos\theta \tag{6.5}$$

and 'the imaginary part of Z is $R\sin\theta$' or

$$Im\ Z = R\sin\theta. \tag{6.6}$$

Note that θ is not unique: adding or subtracting multiples of 2π to θ gives the same complex number.

Complex exponentials

There is yet a third way of expressing a complex number. This is as a complex exponential. It is very important for AC circuits. The properties of exponentials are reviewed in Appendix B. The property of interest here is that

$$\mathrm{e}^{\alpha}\mathrm{e}^{\beta} = \mathrm{e}^{(\alpha+\beta)}. \tag{6.7}$$

Suppose

$$Z_1 = R_1(\cos\theta_1 + \mathrm{j}\sin\theta_1) \qquad \text{and} \qquad Z_2 = R_2(\cos\theta_2 + \mathrm{j}\sin\theta_2).$$

Then

$$\begin{aligned} Z_1Z_2 &= R_1R_2\{(\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + \mathrm{j}(\sin\theta_1\cos\theta_2 + \cos\theta_1\sin\theta_2)\} \\ &= R_1R_2\{\cos(\theta_1+\theta_2) + \mathrm{j}\sin(\theta_1+\theta_2)\}. \end{aligned}$$

Copyright © 2005 IOP Publishing Ltd.

This is just the result obtained by writing

$$Z = R\mathrm{e}^{\mathrm{j}\theta} \equiv R\exp(\mathrm{j}\theta) \tag{6.8}$$

and using (6.7) to multiply Z_1 by Z_2:

$$Z_1 Z_2 = R_1 R_2 \mathrm{e}^{\mathrm{j}(\theta_1+\theta_2)}. \tag{6.9}$$

In words, when two complex numbers Z_1 and Z_2 are multiplied, their phases add and their moduli multiply. The form (6.8) is very useful in relating AC voltage and current.

A particularly important complex number is j itself. It lies on the unit circle, and can be written

$$j = \mathrm{e}^{\mathrm{j}\pi/2}$$

i.e. its phase angle is $\pi/2$. Multiplying any complex number by j therefore has the effect of adding $\pi/2$ to its phase:

$$\mathrm{j}Z_1 = R_1 \mathrm{e}^{\mathrm{j}(\theta_1+\pi/2)}.$$

As an example, multiplying any real number by j turns it into an imaginary number and rotates its phase by $\pi/2$. Likewise,

$$1/\mathrm{j} = -\mathrm{j}^2/\mathrm{j} = -\mathrm{j} = \mathrm{e}^{-\mathrm{j}\pi/2}$$

so dividing any complex number by j has the effect of subtracting $\pi/2$ from its phase:

$$Z_1/\mathrm{j} = R_1 \mathrm{e}^{\mathrm{j}(\theta_1-\pi/2)}.$$

Complex numbers are a generalisation of real numbers. Does the cube root of -1 *or the square root of* j *introduce any new possibility or can they be expressed in terms of complex numbers?*

A final piece of terminology is the **complex conjugate** of a complex number. If $Z = X + \mathrm{j}Y = R\mathrm{e}^{\mathrm{j}\theta}$, its complex conjugate is $Z^* = X - \mathrm{j}Y = R\mathrm{e}^{-\mathrm{j}\theta}$. It is useful because

$$ZZ^* = R^2 = |Z|^2.$$

The complex conjugate of j is $-$j, and $\mathrm{j} \times (-\mathrm{j}) = 1$.

In terms of phase angles, why is $(A + \mathrm{j}B)(A - \mathrm{j}B)$ *real and what is its magnitude?*

Copyright © 2005 IOP Publishing Ltd.

In summary, complex numbers may be expressed in three forms:

$$\boxed{\begin{aligned} Z &= X + \mathrm{j}Y \\ Z &= R(\cos\theta + \mathrm{j}\sin\theta) \\ Z &= R\mathrm{e}^{\mathrm{j}\theta} \end{aligned}}$$

and follow just the same rules of addition and multiplication as real numbers, except for the extra rule that $\mathrm{j}^2 = -1$. Division is simply the converse of multiplication, so that

$$Z_3/Z_2 = Z_1 \qquad \text{where} \qquad R_1 = R_3/R_2 \qquad \text{and} \qquad \theta_1 = \theta_3 - \theta_2. \quad (6.10)$$

Worked examples on complex numbers

Suppose $Z_1 = 3 + 4\mathrm{j}$ and $Z_2 = 5 - 12\mathrm{j}$. They can be rewritten

$$Z_1 = 5\mathrm{e}^{\mathrm{j}\theta_1} \qquad \text{with} \qquad \theta_1 = \tan^{-1}(4/3) = 53.13^\circ$$

$$Z_2 = 13\mathrm{e}^{\mathrm{j}\theta_2} \qquad \text{with} \qquad \theta_2 = \tan^{-1}(-12/5) = -67.38^\circ$$

Then $$Z_3 = Z_2 - Z_1 = (5-3) - (12+4)\mathrm{j} = 2 - 16\mathrm{j}$$

$$Z_4 = Z_1 Z_2 = (15+48) + (20-36)\mathrm{j} = 63 - 16\mathrm{j}$$

$$= 65\mathrm{e}^{\mathrm{j}\theta_4} \qquad \text{with} \qquad \theta_4 = \tan^{-1}(-16/63) = -14.25^\circ$$

and $$Z_5 = Z_2/Z_1 = R_5\mathrm{e}^{\mathrm{j}\theta_5}$$

where $R_5 = 13/5$ and $\theta_5 = \theta_2 - \theta_1 = -120.51^\circ$. Further examples are given at the end of the chapter.

6.2 AC Voltages and Currents

So far, AC problems have been tackled using $V_0 \cos\omega t$ or $V_0 \sin\omega t$. Both give the same answers, since one may be transformed to the other just by redefining $t = 0$. Now

$$\mathrm{e}^{\mathrm{j}\omega t} = \cos\omega t + \mathrm{j}\sin\omega t.$$

It is plausible that the same problems can be done using $V = V_0\mathrm{e}^{\mathrm{j}\omega t}$. Indeed this is so. Then the earlier results may be recovered using

$$Re\ V = V_0\cos\omega t \qquad \text{or} \qquad Im\ V = V_0\sin\omega t.$$

However, there are two great virtues of complex numbers. They actually *simplify* the algebra; and they lead to further insight.

Copyright © 2005 IOP Publishing Ltd.

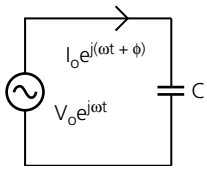


Fig. 6.4. AC applied to a capacitor.

Consider as an example an AC voltage applied to a capacitor, figure 6.4. Voltage and current differ in phase, so we take

$$V = V_0 e^{j\omega t} \qquad \text{and} \qquad I = I_0 e^{j(\omega t+\phi)} = I_0 e^{j\omega t} e^{j\phi}. \tag{6.11}$$

Then

$$V = Q/C$$

$$Q = CV_0 e^{j\omega t}$$

$$I = dQ/dt = j\omega C V_0 e^{j\omega t} = (j\omega C)V. \tag{6.12}$$

In summary,

$$V = IZ \tag{6.13}$$

$$\text{with} \qquad Z = 1/j\omega C = -j/\omega C = (1/\omega C)e^{-j\pi/2}. \tag{6.14}$$

How is this algebra to be interpreted?

From the multiplicative law (6.9) for complex numbers, we can immediately read off from equations (6.13) and (6.14) that:

$$|V| = |Z||I|$$

$$V_0 = (1/\omega C)I_0$$

and

$$\text{Phase}(V) = \text{Phase}(Z) + \text{Phase}(I)$$

$$= -\pi/2 + \text{Phase}(I).$$

These relations summarise what is already known from Chapter 4, but they do so in a very simple form: the reactance of the capacitor is $1/\omega C$ and the current leads voltage by 90° in phase. We can recover the same results as were obtained with sines and cosines by taking the real part of equations (6.11) and (6.12):

$$V = Re(V_0 e^{j\omega t}) = V_0 \cos \omega t$$

$$I = Re[\omega C V_0 e^{j(\omega t+\pi/2)}] = \omega C V_0 \cos(\omega t + \pi/2).$$

Copyright © 2005 IOP Publishing Ltd.

However, the form (6.13) using complex numbers is actually simpler because the complex quantity Z expresses the phase difference; Z is called the **impedance** of the capacitor.

The important points about (6.13) and (6.14) are that

(i) the relation between V and I is linear; this permits us to bring to bear on AC circuits all the pure mathematics of linear algebra; it is the basis of Thevenin's theorem for capacitors (and inductors, see below);

(ii) using complex numbers, the equations express *both magnitude and phase*.

Amplitudes in optics and quantum mechanics are likewise expressed using complex numbers.

6.3 Inductance

For an inductance carrying current $I_0 \cos \omega t$,

$$V_L = L\mathrm{d}I/\mathrm{d}t = -\omega L I_0 \sin \omega t = \omega L I_0 \cos(\omega t + \pi/2). \tag{6.15}$$

Again, the relation may be simplified by writing $I = I_0 \mathrm{e}^{\mathrm{j}\omega t}$, when

$$V_L = L\mathrm{d}I/\mathrm{d}t = \mathrm{j}\omega L I_0 \mathrm{e}^{\mathrm{j}\omega t} = (\mathrm{j}\omega L) I.$$

There is once more a linear relation $V_L = Z_L I$ with

$$Z_L = \mathrm{j}\omega L = (\omega L)\mathrm{e}^{\mathrm{j}\pi/2}. \tag{6.16}$$

This time

$$|V_L| = |Z_L||I| \qquad \text{or} \qquad V_0 = \omega L I_0 \tag{6.17}$$

$$\mathrm{Phase}(V_L) = \mathrm{Phase}(Z_L) + \mathrm{Phase}(I) = \pi/2 + \mathrm{Phase}(I). \tag{6.18}$$

The voltage leads current by 90° in phase. Combining (6.17) and (6.18)

$$V_L = \omega L I_0 \cos(\omega t + \pi/2)$$

which reproduces (6.15).

Copyright © 2005 IOP Publishing Ltd.

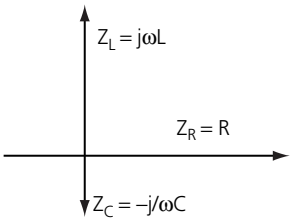


Fig. 6.5. Impedances Z_R, Z_C and Z_L.

6.4 Summary on Impedance

Once you have the idea of using complex numbers, there is very little to memorise. Resistance, capacitance and inductance may be put on the same footing by using complex numbers to write

$$\begin{aligned} V(\text{complex}) &= Z(\text{complex})\ I(\text{complex}) \\ Z(\text{resistance}) &= R \\ Z(\text{capacitance}) &= 1/j\omega C \\ Z(\text{inductance}) &= j\omega L. \end{aligned}$$

In all three cases

$$\begin{aligned} |V| &= |Z||I| \\ \text{phase}(V) &= \text{phase}(Z) + \text{phase}(I). \end{aligned}$$

On the Argand diagram, complex impedances may be represented by the points shown in figure 6.5. This diagram allows us to read off immediately that for an inductance V leads I by a phase angle of 90° and for a capacitance it lags by 90°.

6.5 Impedances in Series

For the series combination shown on figure 6.6:

$$\begin{aligned} V &= V_R + V_L + V_C \\ &= (Z_R + Z_L + Z_C)I \\ &= (R + j\omega L - j/\omega C)I. \end{aligned}$$

Copyright © 2005 IOP Publishing Ltd.

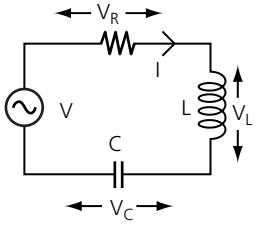


Fig. 6.6. Series combination of R, C and L.

The combined impedance is shown on figure 6.7. Its magnitude is

$$|Z| = \{R^2 + (\omega L - 1/\omega C)^2\}^{1/2}$$

and its phase ψ is given by

$$\tan\psi = (\omega L - 1/\omega C)/R.$$

Voltage leads current by angle ψ and $|V| = |Z||I|$. The impedance is a minimum when $\omega L = 1/\omega C$, i.e. $\omega = 1/\sqrt{LC}$. At this frequency there is a peak in the current of $I = V/R$ and the impedance is purely resistive: $\psi = 0$. This circuit will be studied extensively in Chapter 14.

In a particular problem, we might know I or V or the voltage across one of the elements R, C or L. Getting from one to another involves simple manipulations of magnitudes and phases. Suppose, for sake of example, we know that $V = V_0\cos(\omega t + 30°)$. Then

$$I = I_0\cos(\omega t + 30° - \psi) \qquad \text{where} \qquad I_0 = V_0/|Z|$$

$$V_R = RI = RI_0\cos(\omega t + 30° - \psi)$$

$$V_C = (1/\mathrm{j}\omega C)I = (I_0/\omega C)\cos(\omega t - 60° - \psi)$$

$$V_L = (\mathrm{j}\omega L)I = \omega L I_0\cos(\omega t + 120° - \psi) = \omega L I_0\sin(\omega t + 30° - \psi).$$

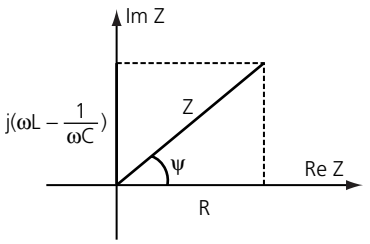


Fig. 6.7. Impedance diagram for the RCL series combination.

Copyright © 2005 IOP Publishing Ltd.

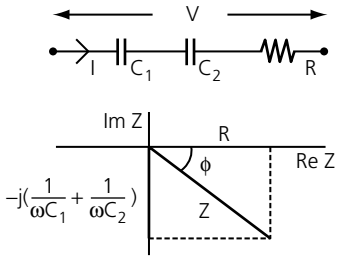


Fig. 6.8. (a) Series combination of C_1, C_2 and R, (b) the corresponding impedance diagram.

In the last line, it is not necessary to remember the trigonometrical relationship between cos and sin. A useful trick is to make use of the fact that

$$e^{j\omega t} = \cos\omega t + j\sin\omega t;$$

then if

$$I \propto \cos(\omega t + 30^\circ - \psi) = Re\{\exp j(\omega t + 30^\circ - \psi)\}$$

$$V_L \propto jI \propto Im\{\exp j(\omega t + 30^\circ - \psi)\} \propto \sin(\omega t + 30^\circ - \psi).$$

A second example is given in figure 6.8. For this circuit

$$V = (R - j/\omega C_1 - j/\omega C_2)I$$

In this case, $V = |Z|I_0\cos(\omega t - \phi)$, where

$$Z = \{R^2 + (1/\omega C_1 + 1/\omega C_2)^2\}^{1/2}$$

$$\tan\phi = (1/\omega C_1 + 1/\omega C_2)/R.$$

This derives the impedance of two capacitors in series.

6.6 Impedances in Parallel

Consider the parallel combination of elements shown in figure 6.9. There is the same voltage V across all three, so

$$I = I_R + I_C + I_L = \frac{V}{R} + \frac{V}{Z_C} + \frac{V}{Z_L} = V\left(\frac{1}{R} + j\omega C + \frac{1}{j\omega L}\right) = V/Z \quad (6.19)$$

$$\frac{1}{Z} = \frac{1}{R} + j(\omega C - \frac{1}{\omega L}). \quad (6.20)$$

Because of the linear relation betwen voltage and current, impedances may be added in series and parallel combinations in just the same way as resistors in

Copyright © 2005 IOP Publishing Ltd.

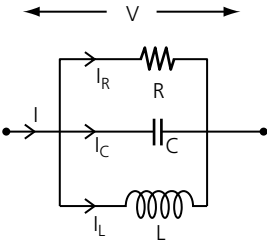


Fig. 6.9. Parallel combination of R, C and L.

Chapter 1. When $\omega C = 1/\omega L$, the second term is zero and $Z \to R$. At this frequency there is a maximum in the impedance and zero phase difference between voltage and current.

There is, however, a final algebraic step required to interpret equations (6.19) and (6.20). They can be used to write

$$|I| = |V||1/Z|$$

$$I_0 = V_0\{(1/R)^2 + (\omega C - 1/\omega L)^2\}^{1/2}$$

$$\text{Phase}(I) = \text{Phase}(V) + \text{Phase}(1/Z) = \text{Phase}(V) + \tan^{-1}\{(\omega C - 1/\omega L)R\}.$$

A resistor has stray capacitance and stray inductance. As $\omega \to \infty$, which dominates: capacitance or inductance?

There may be circumstances where an explicit algebraic form is required for Z itself. From (6.20)

$$Z = \frac{R}{1 + \mathrm{j}(\omega C - 1/\omega L)R}.$$

This expression contains j in the denominator, and if we want to know its real and imaginary parts this may be inconvenient. There is a standard way to rearrange it, by multiplying top and bottom by the complex conjugate of the denominator:

$$Z = \frac{R\{1 - \mathrm{j}(\omega C - 1/\omega L)R\}}{\{1 + \mathrm{j}(\omega C - 1/\omega L)R\}\{1 - \mathrm{j}(\omega C - 1/\omega L)R\}}$$

$$= \frac{R - \mathrm{j}R^2(\omega C - 1/\omega L)}{1 + R^2(\omega C - 1/\omega L)^2}.$$

As a final example, consider the cascade of two CR filters shown in figure 6.10(a). This would be difficult with phasor diagrams. In order to find V_{out}, the elements

Copyright © 2005 IOP Publishing Ltd.

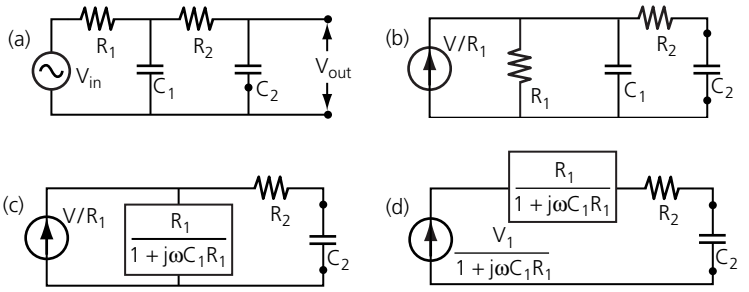


Fig. 6.10. (a) Two filters in cascade, (b)–(d) rearrangement of the circuit.

are rearranged as shown in figures 6.10(b)–(d). The impedances form a potential divider, with the result

$$V_{out} = \frac{V}{1+\mathrm{j}\omega C_1 R_1} \times \frac{1}{\mathrm{j}\omega C_2} \times \frac{1}{\dfrac{1}{\mathrm{j}\omega C_2} + R_2 + \dfrac{R_1}{1+\mathrm{j}\omega C_1 R_1}}$$

$$= \frac{V}{1+\mathrm{j}\omega(C_1R_1 + C_2R_2 + C_2R_1) - \omega^2 C_1 C_2 R_1 R_2}$$

$$\equiv \frac{V}{(1+\mathrm{j}\omega\tau_1)(1+\mathrm{j}\omega\tau_2)}$$

where

$$\tau_1\tau_2 = C_1C_2R_1R_2 = B \qquad \text{say}$$

$$\tau_1 + \tau_2 = \tau_1 + B/\tau_1 = (C_1 + C_2)R_1 + C_2R_2 = 2A \qquad \text{say.}$$

Solving this quadratic equation,

$$\tau_1 = A - (A^2 - B)^{1/2}$$

$$\tau_2 = A + (A^2 - B)^{1/2}.$$

The Bode plot for this circuit is shown in figures 6.11(a) and (b) with corner frequencies at $\omega_1 = 1/\tau_1$ and $\omega_2 = 1/\tau_2$. At high frequencies, the slope is -12 db /octave and the phase lag of the output is π. It should be clear by extrapolation that if n filters are cascaded, the slope at high frequencies is $-6n$ db/octave and the phase lag is $n\pi/2$.

You are advised to practice plenty of examples from the exercises at the end of the chapter, since fluent manipulation of impedances is important throughout all electrical work and electronics.

Copyright © 2005 IOP Publishing Ltd.

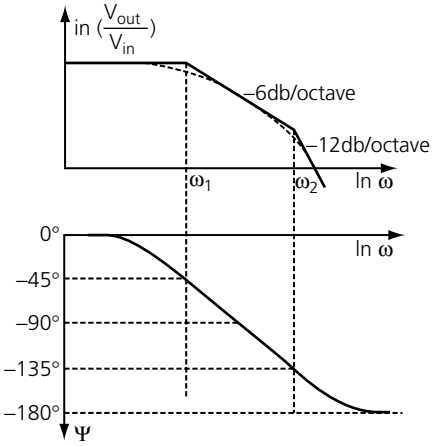


Fig. 6.11. Bode plot for the circuit of Fig. 6.10.

Admittance

Sometimes, it is convenient to work with $Y = 1/Z$ instead of with Z itself. The quantity Y is called admittance. Then, instead of $V = ZI$,

$$I = YV. \tag{6.21}$$

A clear example where this is convenient is in dealing with parallel combinations of elements, as in figure 6.9. In that example,

$$Y = Y_R + Y_C + Y_L$$

$$= \frac{1}{R} + \mathrm{j}\omega C + \frac{1}{\mathrm{j}\omega L} = \frac{1}{R} + \mathrm{j}(\omega C - \frac{1}{\omega L}).$$

The inverse of resistance is called **conductance**, and the official SI unit is the **Siemen**. A common alternative is to call this the **mho** (ohm written backwards).

6.7 Power

Expressions will now be derived for power in terms of complex numbers. Repeating part of the earlier discussion, the power dissipated in an electrical network is given by

$$P = IV = I_0 \cos\omega t \; V_0 \cos(\omega t + \psi) = \tfrac{1}{2} I_0 V_0 \{\cos(2\omega t + \psi) + \cos\psi\}.$$

Averaged over a complete cycle, the first term contributes zero, so the mean power dissipated is

$$\bar{P} = \tfrac{1}{2} I_0 V_0 \cos\psi.$$

Copyright © 2005 IOP Publishing Ltd.

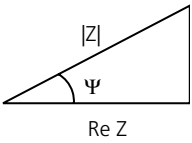


Fig. 6.12. Re Z.

If $V = V_0 e^{j(\omega t+\psi)}$ and $I = I_0 e^{j\omega t}$, then

$$\bar{P} = \tfrac{1}{2} Re(V I^*) = \tfrac{1}{2} Re(V^* I) \tag{6.22}$$

where the star denotes the complex conjugate. From figure 6.12,

$$\cos\psi = \frac{Re(Z)}{|Z|}$$

$$\text{and} \qquad \bar{P} = \tfrac{1}{2} I_0^2 |Z| \frac{Re(Z)}{|Z|} = \tfrac{1}{2} I_0^2 Re(Z). \tag{6.22a}$$

Alternatively,

$$\frac{1}{Z} = |\frac{1}{Z}| e^{-j\psi}$$

$$Re(1/Z) = (1/|Z|)\cos\psi$$

$$\text{and} \qquad \bar{P} = \tfrac{1}{2} V_0^2 Re(1/Z) = \tfrac{1}{2} V_0^2 Re(Y). \tag{6.22b}$$

All of equations (6.22) are simple and obvious generalisations of the corresponding results for resistors.

6.8 Bridges

The familiar Wheatstone bridge is generalised in the arrangement shown in figure 6.13, where impedances $Z_1 \ldots Z_4$ replace resistances. The out-of-balance

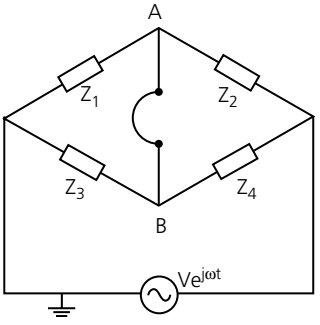


Fig. 6.13. AC Wheatstone bridge.

Copyright © 2005 IOP Publishing Ltd.

current between A and B may be measured by headphones or by a multimeter on its AC range. The condition for zero current in this arm is that $V_A = V_B$. If the current in the arm AB is zero,

$$V_A = V\mathrm{e}^{\mathrm{j}\omega t}\frac{Z_1}{Z_1 + Z_2}$$

$$V_B = V\mathrm{e}^{\mathrm{j}\omega t}\frac{Z_3}{Z_3 + Z_4}$$

so the condition for balance is that

$$\frac{Z_3 + Z_4}{Z_3} = \frac{Z_1 + Z_2}{Z_1}$$

$$Z_4/Z_3 = Z_2/Z_1 \tag{6.23}$$

$$Z_1 Z_4 = Z_2 Z_3. \tag{6.23a}$$

These equations requires that both real and imaginary parts of (6.23) balance.

Many types of bridge have been devised to measure capacitances and inductances. One worked example will be given here and others in the exercises. A practical warning is that stray capacitance between points in the circuit and stray inductance in the resistors may present problems, particularly at high frequency.

If the generator does not give pure sine waves but contains harmonics, does it alter the balance point of the bridge? If so, what can you do about it?

Worked example

For the Owen bridge of figure 6.14, equation (6.23a) gives

$$R_1(R_4 + \frac{1}{\mathrm{j}\omega C_4}) = (R_2 + \mathrm{j}\omega L_2)\frac{1}{\mathrm{j}\omega C_3}.$$

This requires

$$R_1 R_4 = L_2/C_3$$

$$R_1/C_4 = R_2/C_3.$$

The variable resistors R_2 and R_4 may be adjusted independently to satisfy these two conditions. It is necessary to adjust R_2 and R_4 successively to achieve balance. The balance in this particular bridge is independent of ω; this is not always the case.

Copyright © 2005 IOP Publishing Ltd.

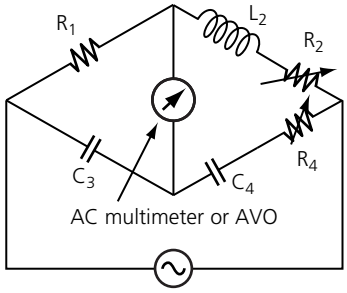


Fig. 6.14. The Owen bridge.

6.9 Exercises

Examples 1–5 revise complex numbers; applications to circuits follow.

1. *Representation of complex numbers*. Draw real and imaginary axes on graph paper and plot the following complex numbers. Convert each one to polar form $re^{j\phi}$. (a) $3+j4$, (b) $-3+j3$, (c) $-8-6j$, (d) -1, (e) $-2j$. (Ans: $5e^{j53.13^\circ}$, $3\sqrt{2}e^{135^\circ}$, $10e^{j216.87^\circ}$, $1e^{j180^\circ}$, $2e^{270^\circ}$.)

2. *Products of complex numbers*. Calculate the products of the following pairs of complex numbers. As a check, convert them into polar form, and again compute their product. (a) $(3+j4)(-3+j3)$; (b) $(-3+j3)(-8-6j)$; (c) $(-8-6j)(-2j)$; (d) $(x-jy)(x+jy)$. (Ans: $-21-j3 = 15\sqrt{2}e^{j188.13^\circ}$; $42-6j = 30\sqrt{2}e^{-j8.13^\circ}$; $-12+16j = 20e^{j126.87^\circ}$; x^2+y^2.)

3. *Quotients of complex numbers*. Find the quotient in expressions (a)–(d) by multiplying both numerator and denominator by the complex conjugate of the denominator. As a check, convert the numbers to polar form, and evaluate the quotient in this form. (a) $j7/(3-j3)$; (b) $(3+j4)/(-3+j3)$; (c) $(-3+j3)/(3+j4)$; (d) $(3+j3)/(1-j1)$. (Ans: $7(j-1)/6 = 7e^{j135^\circ}/3\sqrt{2}$; $(1-7j)/6 = 5e^{-j81.87^\circ}/3\sqrt{2}$; $3(1+7j)/25 = 3\sqrt{2}e^{j81.87^\circ}/5$; $3j = 3e^{j90^\circ}$.)

4. $e^{j\phi}$. Express each complex number in cartesian form and plot graphically. (a) $3e^{-j\pi/4}$, (b) $4e^{j\pi/2}$, (c) $5e^{-j2\pi/3}$, (d) $6e^{j4\pi/3}$. (Ans: $3(1-j)/\sqrt{2}$, $4j$, $-5(1+\sqrt{3}j)/2$, $3(-1-\sqrt{3}j)$.)

5. $e^{j(\omega t+\phi)}$. A point moves in the complex plane with time as $e^{jt/2}$. Plot its movement at $t = 0,1,2 \ldots 10$. What happens after $t = 4\pi$? Plot a graph against t of $Re(e^{jt/2})$ and $Re\{e^{j(t/2+\pi/2)}\}$. Satisfy yourself that $d(e^{jt/2})/dt = (j/2)e^{jt/2}$. (Ans: repeats after $t = 4\pi$.)

6. An inductance of 4 mH passes a current $I = 5\cos 500t$ mA. What is the voltage across it? Show that $I = 5Re(e^{j500t})$ mA and find the voltage in terms of complex exponentials. Repeat this problem for a capacitance

Copyright © 2005 IOP Publishing Ltd.

of 20 μF replacing L. (Ans: Re $(10\mathrm{j}\mathrm{e}^{\mathrm{j}500t}) = -10\sin 500t$ mV; Re $(0.5\mathrm{e}^{\mathrm{j}500t}/\mathrm{j}) = 0.5\sin 500t$ V.)

7. *Addition and subtraction of complex amplitudes*. (a) Represent the two currents $I_1 = 2\sin(\omega t - \pi/4)$ and $I_2 = \sin(\omega t + \pi/2)$ on the complex plane at $t = 0$ and hence find $I_1 + I_2$ for any t; (b) find the difference $V_1 - V_2$ where $V_1 = 5\cos(\omega t + 50°)$ and $V_2 = 3\cos(\omega t + 150°)$. (Ans: $1.47\sin(\omega t - 16.32°)$; $6.26\cos(\omega t + 21.85°)$.)

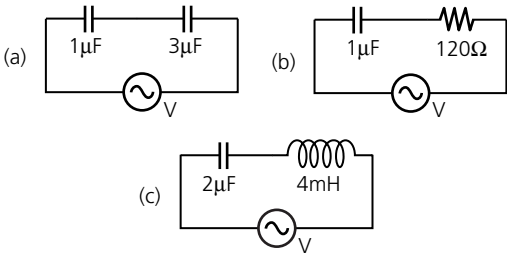


Fig. 6.15.

8. *Complex impedance*. If in figures 6.15(a)–(c) $V = 10\mathrm{e}^{\mathrm{j}\omega t}$ mV, where $\omega = 10^4$ rad/s find the magnitude and phase of the current and the voltage across each of the six components. (Ans: (a) $0.075\mathrm{e}^{\mathrm{j}(\omega t+\pi/2)}$ mA, $7.5\mathrm{e}^{\mathrm{j}\omega t}$ mV, $2.5\mathrm{e}^{\mathrm{j}\omega t}$ mV; (b) $64\mathrm{e}^{\mathrm{j}(\omega t+39.81°)}$ μA, $6.4\mathrm{e}^{\mathrm{j}(\omega t-50.19°)}$ mV, $7.68\mathrm{e}^{\mathrm{j}(\omega t+39.81°)}$ mV; (c) $\mathrm{e}^{\mathrm{j}(\omega t+\pi/2)}$ mA, $50\mathrm{e}^{\mathrm{j}\omega t}$ mV, $-40\mathrm{e}^{\mathrm{j}\omega t}$ mV.)

9. If the components in each of figures 6.15(a)–(c) are rearranged in parallel, what current flows in each component, and what is the total current in each circuit? (Ans: (a) $100\mathrm{e}^{\mathrm{j}(\omega t+\pi/2)}$ μA, $300\mathrm{e}^{\mathrm{j}(\omega t+\pi/2)}$ μA, $400\mathrm{e}^{\mathrm{j}(\omega t+\pi/2)}$ μA; (b) $100\mathrm{e}^{\mathrm{j}(\omega t+\pi/2)}$ μA, $83.3\mathrm{e}^{\mathrm{j}\omega t}$ μA, $130.17 \times \mathrm{e}^{\mathrm{j}(\omega t+50.19°)}$ μA; (c) $0.25\mathrm{e}^{\mathrm{j}(\omega t-\pi/2)}$ mA, $0.05\mathrm{e}^{\mathrm{j}(\omega t-\pi/2)}$ mA, $-0.2\mathrm{e}^{\mathrm{j}(\omega t-\pi/2)}$ mA.)

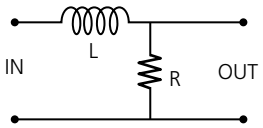


Fig. 6.16.

10. A low pass filter consists of a capacitance C and resistance R in series. Show that the output voltage and input voltage are related by $V_{out}/V_{in} = 1/(1 + \mathrm{j}\omega CR)$. The same resistance is used in series with an inductance L to make a low pass filter of identical frequency dependence. Draw the circuit and find the value of L. (Ans: figure 6.16, $L = CR^2$.)

Copyright © 2005 IOP Publishing Ltd.

11. (QMC, 1975). A series LCR circuit with $L = 15$ mH has an instantaneous voltage $V = 45\sin(4000t + 30°)$ mV and an instantaneous current $I = 0.5\sin(4000t)$ mA. Find the values of R and C. At what frequency would the voltage and current always be in phase? (Ans: $45\sqrt{3}\Omega$, 16.7 μF, $\omega = 2000$ rad/s, $f = 1000/\pi$ Hz.)

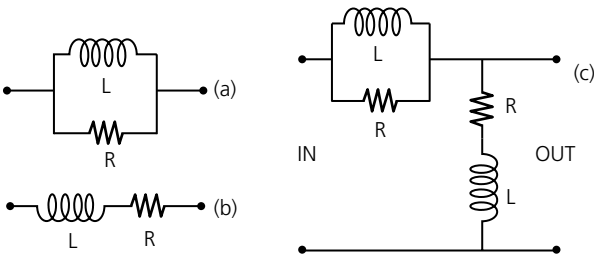


Fig. 6.17.

12. Find expressions for the impedances of the networks in figures 6.17(a) and (b). At what frequency does circuit 6.17(c) produce zero phase shift? (Ans: (a) $Z = \mathrm{j}\omega LR/(R + \mathrm{j}\omega L)$, (b) $Z = R + \mathrm{j}\omega L$; $\omega = R/L$.)

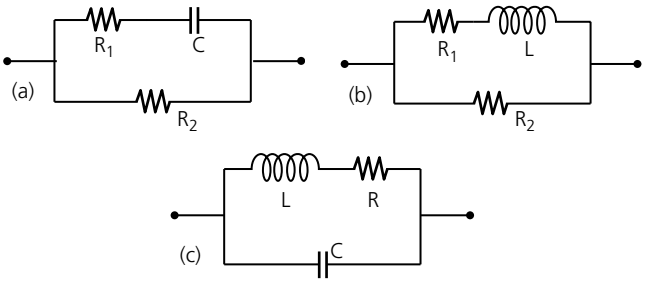


Fig. 6.18.

13. Find the impedances of the networks in figures 6.18(a), (b) and (c). (Ans: $R_2(1 + \mathrm{j}\omega CR_1)/\{1 + \mathrm{j}\omega C(R_1 + R_2)\}$, $R_2(R_1 + \mathrm{j}\omega L)/(R_1 + R_2 + \mathrm{j}\omega L)$, $(R + \mathrm{j}\omega L)/\{(1 - \omega^2 CL) + \mathrm{j}\omega CR\}$.)

14. (QMC). What is the power factor in circuits used at angular frequency ω consisting of (a) an ideal capacitor, (b) an inductance L in series with a resistance $R = \omega L$, * (c) the elements (a) and (b) in parallel? (Ans: (a) 0, (b) $1/\sqrt{2}$, (c) $\{2 - 4\omega^2 LC + 4\omega^4 L^2 C^2\}^{-\frac{1}{2}}$.)

15. Find the conditions for zero signal in the detector D of Maxwell's L/C bridge, figure 6.19. (Ans: $S/R = Q/P$ and $L = CQR$.)

16. The Maxwell self-inductance bridge of figure 6.20 is used to find the resistance R_4 and self-inductance L_4 of an inductor by varying L_3 and

Copyright © 2005 IOP Publishing Ltd.

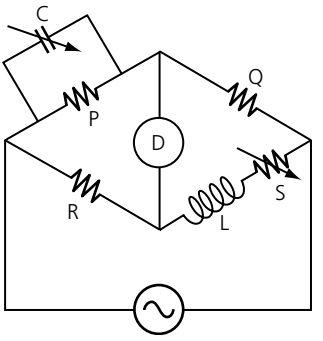


Fig. 6.19.

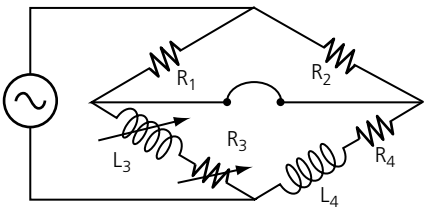


Fig. 6.20.

R_3. Find the balance conditions. (Ans: $R_4 = R_2R_3/R_1$ and $L_4 = R_2L_3/R_1$.)

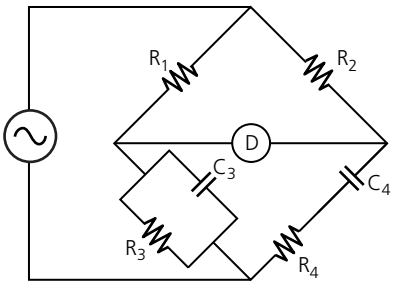


Fig. 6.21.

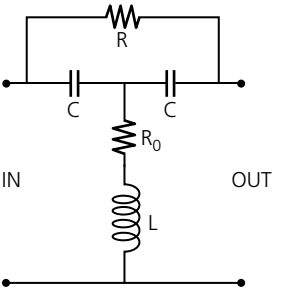


Fig. 6.22.

17. Find the balance conditions for the Wien bridge of figure 6.21. (Ans: $\omega^2 = 1/C_3C_4R_3R_4$ and $(C_3/C_4) + (R_4/R_3) = R_2/R_1$.)

18*. (QEC). Find the conditions for zero output from the network of figure 6.22. (Ans: $\omega^2CLR = 2R_0$ and $RR_0 + 2L/C = (1/\omega^2C^2)$.)

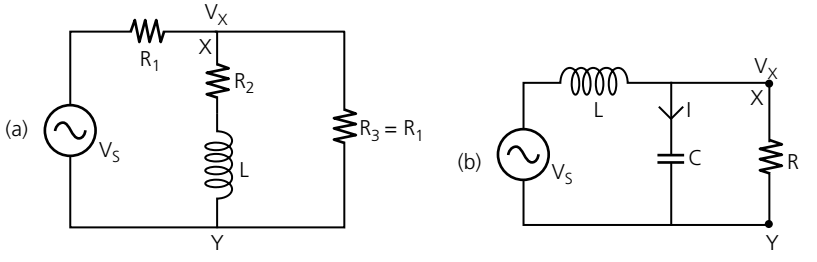


Fig. 6.23.

Copyright © 2005 IOP Publishing Ltd.

19. *Node voltage and Thevenin's theorem.* Find V_X in figures 6.23(a) and (b) by using Kirchhoff's current law for that node. What is the Thevenin equivalent circuit across the points XY in both cases? Use this to check your evaluation of V_X. (Ans: (a) $\frac{1}{2}V_S/\{1+\frac{1}{2}R_1/(R_2+\mathrm{j}\omega L)\}$, $V_{EQ}=\frac{1}{2}V_S$, $Z_{EQ}=\frac{1}{2}R_1$; (b) $V_S/\{1-\omega^2CL+\mathrm{j}\omega L/R\}$, $V_{EQ}=V_S/(1-\omega^2CL)$, $Z_{EQ}=\mathrm{j}\omega L/(1-\omega^2CL)$.)

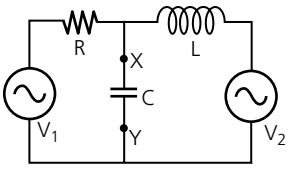


Fig. 6.24.

20**. *Superposition and Thevenin's theorem.* Find the current through the capacitor of figure 6.24 originating separately from V_1 and V_2. Check by finding the Thevenin equivalent circuit across terminals XY and hence the current I. (Ans: $I_1=-V_1\omega^2CL/\{R(1-\omega^2CL)+\mathrm{j}\omega L\}$, $I_2=V_2\mathrm{j}\omega CR/\{R(1-\omega^2CL)+\mathrm{j}\omega L\}$, $V_{EQ}=(V_2R+V_1\mathrm{j}\omega L)/(R+\mathrm{j}\omega L)$, $Z_{EQ}=\mathrm{j}\omega LR/(R+\mathrm{j}\omega L)$.)

21. A generator with output impedance $R_1+\mathrm{j}X_1$ is connected directly across a load $R_2+\mathrm{j}X_2$. Show that the maximum power is transferred to the load when $R_1=R_2$ and $X_1=-X_2$. Can you interpret this last result physically?

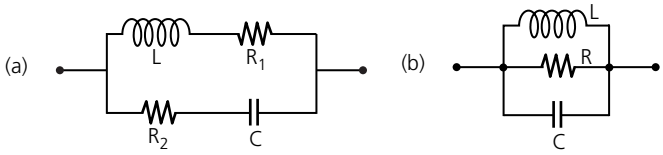


Fig. 6.25.

22. Find the admittance of the networks in figures 6.25(a) and (b). (Ans: $\{1-\omega^2CL+\mathrm{j}\omega C(R_1+R_2)\}/\{(R_1+\mathrm{j}\omega L)(1+\mathrm{j}\omega CR_2)\}$ and $(1/R)+\mathrm{j}(\omega^2CL-1)/\omega L$.)

Copyright © 2005 IOP Publishing Ltd.

7

Operational Amplifiers and Negative Feedback

7.1 Introduction

So far, we have mostly considered passive elements (resistors, capacitors and inductors), which simply transmit currents and voltages. Amplifiers involve active devices – transistors. The amplifier is a central feature of all electronic circuits. Sometimes it may be hidden from view inside the circuit and may not be immediately apparent in the relation between input and output, but it is always there. Even in digital circuitry where the output voltage is often the same as at the input, the circuit amplifies current.

There are certain general principles which allow circuits to be designed so that they are insensitive to the precise details of the amplifier. In consequence, circuits designed 40 or 60 years ago for valves can be redesigned with changes of detail to work with bipolar transistors or FET's or any other amplifier. As each new device is invented, the principles of circuit design do not change in an essential way though details do.

This chapter will be concerned with those principles. For the moment the properties of individual transistors can be sidestepped. It is fortunate that manufacturers have done much of the detailed hard work by marketing cleverly designed integrated circuits (containing many transistors) which act as nearly ideal amplifiers. They are conservatively designed to avoid common pitfalls; as long as we do not want to stretch performance to the limit they are a great convenience. Those of you who aspire to the ultimate in performance (or to design actual integrated circuits) will later need to study the idiosyncracies of individual transistors.

Consider initially a voltage amplifier for which

$$\boxed{\text{voltage gain, } G = \frac{\text{output voltage}}{\text{input voltage}}.} \tag{7.1}$$

Ideally, G is large and independent of input voltage, current and frequency. The ideal voltage amplifier has large input resistance, so that it puts a negligible load

Copyright © 2005 IOP Publishing Ltd.

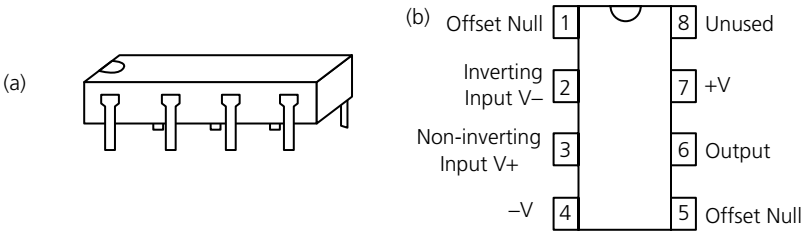


Fig. 7.1. (a) An operational amplifier as a 'chip', (b) the pin layout of the μA741C.

on the previous circuit. It has a low output resistance, so that most of the output voltage is delivered to the subsequent circuit.

Opamps

In the laboratory, you can experiment with individual **operational amplifiers** or 'opamps' which satisfy most of these criteria. They are the building blocks which go into modern electronic 'chips', which may contain thousands of individual amplifiers. The design of complicated integrated circuits is an intricate art today and far beyond the scope of this book. As an introduction to the principles, it is essential to understand how operational amplifiers are used.

The $\mu A741C$ will be adopted for illustrative purposes. Its layout is shown in figure 7.1. Connections are made via the legs, which can be pushed into a breadboard. The legs are easily bent, and you need to ease the chip out of a circuit board carefully using pliers. Power supplies, typically +15 and 15 V are connected at pins 7 and 4 labelled $+V$ and $-V$. Ultimately, the gain of the amplifier arises by drawing currents from these supplies. For reasons which will soon become apparent, all operational amplifiers are actually differential amplifiers having two inputs V_+ and V_- (at pins 3 and 2 in this case) and

$$V_{out} = G(\omega)(V_+ - V_-). \tag{7.2}$$

Pin 3 is called the non-inverting input and pin 2 the inverting input: if $V_+ = 0$, $V_{out} = -G(\omega)V_-$. The remaining pins are used for fine tuning of the performance.

Figure 7.2 shows a schematic representation of the amplifier. Its essential characteristics are:

(a) input impedance $r_{in} \sim 1$ MΩ; (for some other types of operational amplifier this can be as high as 10^{12} Ω);
(b) output impedance $r_{out} \sim 75$ Ω; small letters r_{in} and r_{out} are used since these are usually resistances of active transistors rather than resistors;
(c) temperature stability $\mathrm{d}G/G \sim 10^{-5}/^{\circ}$C;
(d) supply voltages $+V$ and $-V$ which may be varied from 3 to 18 V; they need not be the same magnitude;
(e) input and output protected against short circuits and against voltages up to $\pm$15 V;

Copyright © 2005 IOP Publishing Ltd.

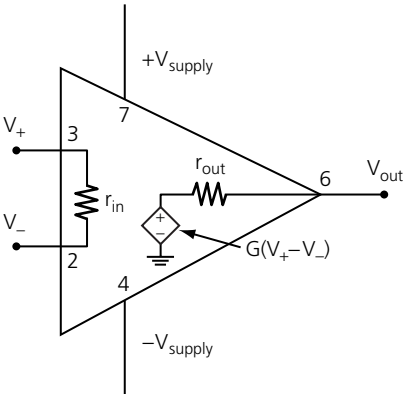


Fig. 7.2. (a) Schematic of the amplifier.

(f) output currents up to 20 mA;
(g) a gain G up to 2×10^5 at low frequencies; however, the output voltage reaches a limit when it gets within about 1 V of the power supply voltages. The gain varies with frequency as shown in figure 7.3. This design is desirable as a safeguard against simple amplifiers breaking into oscillation.

7.2 Series Voltage Feedback

Such an amplifier is hardly ever used in isolation. Almost invariably it is built into a **feedback loop**. Feedback can be used to control the performance of the amplifier

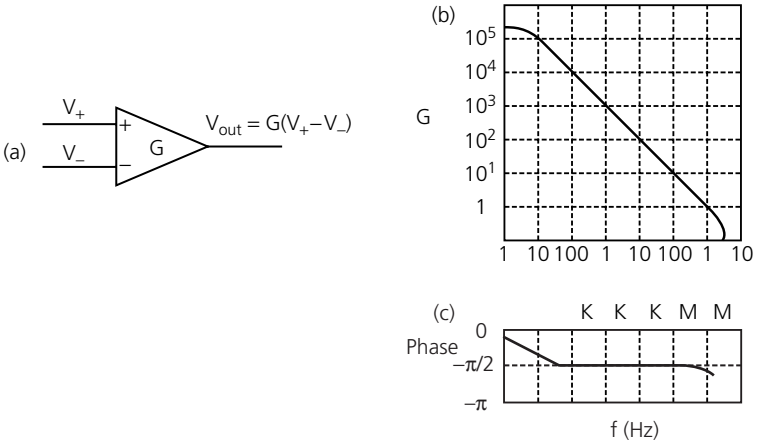


Fig. 7.3. (a) Symbol for a differential amplifier, (b) G(f) of the μA741C, (c) phase difference between output and input.

Copyright © 2005 IOP Publishing Ltd.

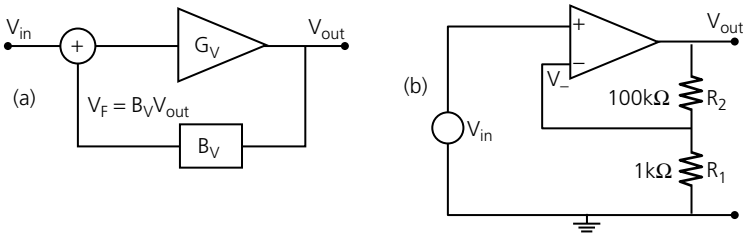


Fig. 7.4. (a) A general feedback loop, (b) negative feedback.

and is fundamental to the design of electronic circuits. It is also fundamental to Control Systems.

Figure 7.4(a) shows schematically the general case where a feedback voltage $B_v V_{out}$ is derived from the output voltage and fed back to the input of the amplifier. There it is summed with the input voltage V_{in}. Then

$$V_{out} = G_v(V_{in} + V_F) = G_v(V_{in} + B_v V_{out})$$

providing the output voltage does not saturate; the amplification A of the whole circuit is

$$\boxed{A = \frac{V_{out}}{V_{in}} = \frac{G_v}{1 - G_v B_v}.} \tag{7.3}$$

This formula plays a fundamental role in the design of electronic circuits. Note the distinction between G_v, the voltage gain of the operational amplifier itself and A, the amplification of the whole circuit between input and output voltages. The quantity G_v is called the **open-loop gain**, i.e. the gain with no feedback; A is called the **closed-loop gain**.

The feedback fraction B_v may be positive or negative. **Positive feedback** ($G_v B_v > 0$) is used in digital circuitry. **Negative feedback** ($G_v B_v < 0$) will be discussed at length in this chapter. One simple way of achieving it is shown in figure 7.4(b), where the feedback is applied from a potential divider to the negative terminal of a differential amplifier. In this case

$$G_v = G(\omega)$$

$$V_F = -\{R_1/(R_1 + R_2)\}V_{out}$$

$$B_v = -R_1/(R_1 + R_2) = -B \qquad \text{say.}$$

Then

$$\boxed{A = \frac{G(\omega)}{1 + BG(\omega)} \qquad \text{for negative feedback.}} \tag{7.3a}$$

You will meet both forms (7.3) and (7.3a) in the literature and the difference in sign in the denominator is a source of confusion. The former covers in general the

Copyright © 2005 IOP Publishing Ltd.

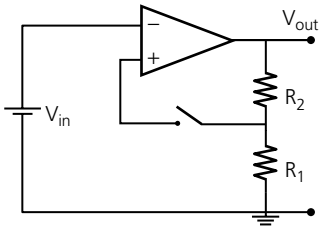


Fig. 7.5. Positive feedback.

use of positive or negative feedback. The latter conveniently takes account of the minus sign when B_v is negative.

For negative feedback, there is an important simplification if $BG \gg 1$. Then the 1 in the denominator can be neglected and

$$\boxed{A \simeq 1/B.} \tag{7.4}$$

(This result assumes the output does not saturate.) The great virtue of this arrangement is that amplification is independent of G, which can vary from one batch of amplifiers to another. If $G = 10^5$ and $B = 0.01$, then $A \simeq 100$. The exact solution (7.3a) gives $A = 99.90$ if $G = 10^5$ and 99.95 if $G = 2 \times 10^5$, showing that a change of G by a factor 2 alters the circuit amplification A by only 0.05 %. Since B is derived from a pair of resistors, it is very stable and independent of frequency. Achieving this degree of stability is otherwise virtually impossible. The use of negative feedback makes circuit performance insensitive to the amplifier. It is a general characteristic of negative feedback that it enhances stability. We shall see later that it has other virtues too.

Everyday examples

Negative feedback is commonplace in everyday life. If you argue a point with someone, he or she is likely to put the reverse point of view. Checks and balances in politics and the judiciary are important for stability. In the home, a thermostat measures room temperature and cuts off the heating when the temperature gets too high. When you drive a car you apply feedback to stabilise speed and direction. In Nature, feedback stabilises the population via the food supply. In every case, the signal fed back to the input depends on the output, making a closed loop.

Positive feedback

Positive feedback on the other hand leads to instability or saturation of the amplifier. You might think that we can still make the approximation $G_v B_v \gg 1$ and arrive at the result $A = -1/B_v$ again. However, what happens in this case is that the amplifier saturates. Suppose that V_{in} of figure 7.5 is a very small signal, so small that the output of the amplifier does not saturate. If the switch is

Copyright © 2005 IOP Publishing Ltd.

closed and feedback is applied to the positive input of the amplifier, it reinforces V_{out} and drives it to the limit of what the amplifier will produce. When the amplifier saturates, G_v falls until $G_v B_v$ is just less than 1. Equation (7.3) is then satisfied with V_{out} equal to whatever happens to be the voltage limit of the amplifier; G_v takes up an appropriate value less than its unsaturated value. If V_{in} is reversed in sign, the same feedback loop drives the amplifier to saturation with the other polarity. In either case, the output of the amplifier swings to its extreme positive or negative limit and is insensitive to the precise value of the input. These two saturated states are useful for representing logical 0 and 1 in digital circuitry.

In economics, positive feedback is dangerous because it leads to violent swings. War is also like this with the population united in a common aim. Dictators surround themselves with 'Yes Men' to achieve positive feedback within their empires; changes of dictator are frequently violent.

Let us return to negative feedback and equations (7.3a) and (7.4). You might be tempted to make B very small so as to achieve a large amplification $1/B$. However, remember the approximation $GB \gg 1$. The gain falls at high frequency according to figure 7.3, and the approximation will fail above frequencies where $GB = 1$. This limits the bandwidth of the circuit. On the straight part of the Bode plot, $G = G_0/\omega$, where $G_0 = 6 \times 10^6$ rad/s for the $\mu A741C$. Suppose an audio amplifier is required working up to $\omega = 10^5$ rad/s. If GB is to be 1 at this frequency, $B = \omega/G_0 = 1/60$, and the amplification of a single stage is limited to 60. If more amplification is required, the remedy is to cascade several stages of amplification. The upper limit of frequency at which amplification is possible with the μA741C is 6×10^6 rad/s where $G = 1$. If an amplifier is required for higher frequencies, this chip has insufficient bandwidth and a superior operational amplifier must be used.

It is instructive to put together the circuit of figure 7.4(b) and examine the amplification with a DC input then AC signals as a function of frequency. Input and output voltages may be displayed on two traces of an oscilloscope. Try varying the magnitude of V_{in} until the output saturates. Then vary positive and negative supply voltages and observe their effect on the levels at which the output limits. Try varying the ratio of R_2 to R_1.

A practical point is that no pin of the operational amplifier is earthed, see figure 7.1(b). How does it know where Earth potential is with respect to the supply voltages $+V$ and $-V$? The answer is via the load applied across V_{out}, e.g. an oscilloscope input resistance if you are observing the waveform. It is important to earth the supply voltages at the same point as the load (except when you specifically wish to introduce a bias voltage). If the supply voltages have a floating earth, the circuit will not function properly.

A second practical point is that the amplifier does actually draw small DC bias currents for transistor operation through the negative and positive input terminals V_+ and V_-. You must not cut off these bias currents by capacitors; if you do, the amplifier will not work: the intention is that it should be DC coupled.

Copyright © 2005 IOP Publishing Ltd.

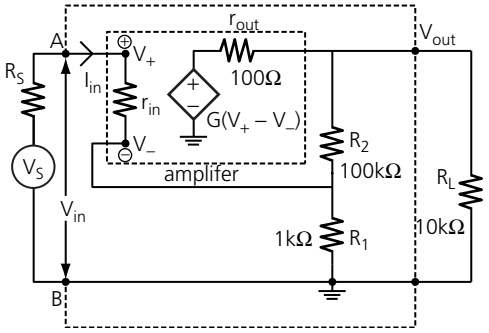


Fig. 7.6. Voltage feedback showing greater detail.

7.3* Appoximations in Voltage Feedback

The alert reader may have noticed that there were some approximations in reaching equation (7.3a). We pause to sketch these. With the component values chosen for figure 7.4(b) they are actually very good, but the unwary might choose much higher or lower values of the feedback resistors R_1 and R_2 and run into trouble.

Figure 7.6 shows the circuit of figure 7.4 drawn out in more detail. It includes a load resistor R_L. There are two voltage generators: the source V_S which supplies the circuit and the generator $G(V_+ - V_-)$ at the output of the operational amplifier. The currents in the circuit are the superposition of those from these two voltage sources, and the voltages in the circuit follow from the sum of these currents; deriving V_- from them is an alternative (and more precise) way of deriving the voltage gain.

The first approximation was that $V_{out} = G(V_+ - V_-)$, neglecting the voltage drop in r_{out}. This is a good approximation if $R_L \gg r_{out}$. However the resistance from the output of the amplifier to earth is really the parallel combination of r_{out}, R_L and $R_1 + R_2$; this is a warning not to make R_L much smaller than r_{out}. Also R_1 is in parallel with $r_{in} + R_S$ through the input of the amplifier. This is not a serious consideration since $r_{in} \simeq 10^6\ \Omega$. Likewise, R_S should not be made larger than r_{in} or the input signal will be attenuated before it gets to the amplifier. A final detail is that V_{in} actually makes a direct contribution to V_{out} via the current it drives through R_2. But again this is a negligible effect if resistors are chosen sensibly. The moral is that it is worth drawing out the full circuit diagram, figure 7.6, and making sure that the usual approximations are adequate; if not, they are easily corrected.

7.4 Shunt Feedback

Figure 7.7 shows an alternative feedback arrangement which is used more commonly. The output is connected back to the negative input of the operational amplifier via the feedback resistor R_F. This supplies a feedback current to the

Copyright © 2005 IOP Publishing Ltd.

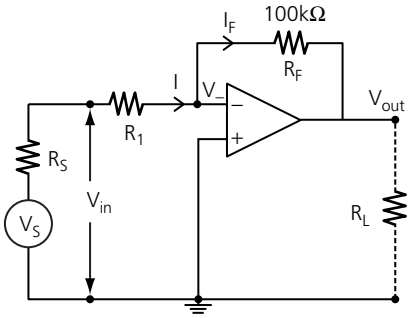


Fig. 7.7. Shunt feedback, $V_{out} = -(R_F/R_1)V_{in}$.

amplifier. For this reason, the configuration is sometimes called **negative current feedback**. An alternative name, used here, is **shunt feedback**. This is because R_F appears across the operational amplifier, between input and output.

There is a simple but powerful way of deriving a close approximation to the amplification V_{out}/V_{in}. It will be demonstrated here and used in many exercises at the end of the chapter. From equation (7.2), $V_+ - V_- = V_{out}/G$. If G is large,

$$\boxed{V_+ \simeq V_-.} \tag{7.5}$$

In figure 7.7, where V_+ is earthed, this ensures that the negative terminal is very close to Earth potential, despite the fact that there is no direct electrical connection to Earth. The negative terminal is said to be a **Virtual Earth**. Then the current through resistor R_1 is $I \simeq V_{in}/R_1$. Very little of this current flows into the operational amplifier for two reasons. Firstly, $V_+ - V_-$ is very small; secondly, r_{in} is usually large compared with R_F. The current I_F through the feedback resistor R_F is therefore very close to I. Because $V_- \simeq 0$,

$$V_{out} \simeq -R_F I = -R_F V_{in}/R_1$$

so the amplification A is given by

$$\boxed{A = V_{out}/V_{in} \simeq -R_F/R_1.} \tag{7.6}$$

The arrangement in figure 7.7 acts rather like a lever pivoting about zero at the negative terminal. It gives inversion and an amplification independent of G; the result looks like that for voltage feedback, equation (7.4), except for a sign change and $B = R_1/R_F$. The sign change arises because V_{in} is attached to the negative terminal of the amplifier. Again, the formula for the amplification assumes that the input voltage is small enough that the output does not saturate.

Although the algebra given above is very simple, it does hide one vital point. There is actually a small difference between V_+ and V_- and it is this difference which drives the amplifier. The algebra will be repeated keeping this small effect,

Copyright © 2005 IOP Publishing Ltd.

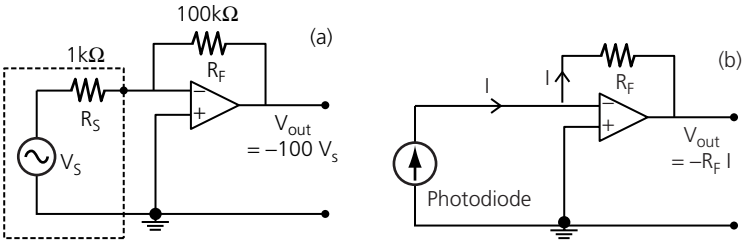


Fig. 7.8. (a) Amplifying by 100, (b) current to voltage conversion.

but retaining the approximation that no current flows into the operational amplifier. Then

$$V_- = -V_{out}/G$$

$$I = (V_{in} - V_-)/R_1$$

$$V_{out} = V_- - I R_F = V_- \left(1 + \frac{R_F}{R_1}\right) - V_{in}\frac{R_F}{R_1}$$

$$= -\frac{V_{out}}{G}\left(1 + \frac{R_F}{R_1}\right) - V_{in}\frac{R_F}{R_1}.$$

Finally

$$V_{out}\left[1 + \frac{1}{G}\left(1 + \frac{R_F}{R_1}\right)\right] = -V_{in}\frac{R_F}{R_1}.$$

The effect of the small difference between V_+ and V_- is to introduce an extra term of order $1/G$ on the left hand side. At low frequencies this term is negligible because G is so large; it simplifies the algebra greatly to omit it.

The current I flows towards the input of the amplifier if V_{in} is positive; I_F flows away from it. The input current I is almost exactly balanced by the feedback current I_F. This is the origin of the term **negative current feedback**.

Shunt feedback is used very widely. Suppose, for example, we want to amplify by 100 the voltage from a previous circuit having an output impedance of 1 kΩ. This output impedance can play the role of R_1 (providing its value is sufficiently stable) and then R_F is chosen to be 100 $k\Omega$, figure 7.8(a). The resulting amplification is independent of the choice of operational amplifier, providing that it has $G \gg 100$ and adequate bandwidth.

This simple arrangement is also a convenient way of transforming a current to a voltage. Suppose, for example, a current source such as a photodiode is available. It supplies an input current I in figure 7.8(b) and $V_{out} = -R_F I$.

Copyright © 2005 IOP Publishing Ltd.

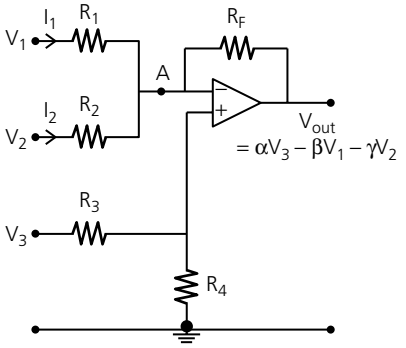


Fig. 7.9. The analogue adder.

7.5 The Analogue Adder

The previous circuit produced an output voltage proportional to input voltage. Figure 7.9 shows a circuit which generalises this by adding and subtracting input signals; the output voltage is $\alpha V_3 - \beta V_1 - \gamma V_2$ where coefficients α, β and γ depend only on resistor values.

To see in outline how the circuit works, remember that (a) $V_+ \simeq V_-$, (b) the current flowing into the operational amplifier itself is negligibly small (at low frequencies). Then V_+, the voltage at the positive terminal of the operational amplifier, is determined from V_3 by the potential divider R_3 and R_4. Next $V_- \simeq V_+$. So the currents through R_1 and R_2 are proportional to $(V_1 - V_+)$ and $(V_2 - V_+)$ respectively; these currents add and flow through R_F, so V_{out} depends linearly on V_1, V_2 and V_+ (hence V_3). This arrangement can be generalised to any number of inputs.

It is simple algebra to find the coefficients α, β and γ:

$$
\begin{aligned}
V_+ &= V_3 R_4/(R_3 + R_4) \\
V_- &\simeq V_+ \\
V_{out} &\simeq V_+ - R_F(I_1 + I_2) \\
&\simeq V_+ - R_F\left(\frac{V_1 - V_+}{R_1}\right) - R_F\left(\frac{V_2 - V_+}{R_2}\right) \\
&= V_3 \frac{R_4}{R_3 + R_4}\left(1 + \frac{R_F}{R_1} + \frac{R_F}{R_2}\right) - \frac{R_F}{R_1}V_1 - \frac{R_F}{R_2}V_2 \qquad (7.7) \\
\alpha &= \frac{R_4}{R_3 + R_4}\left(1 + \frac{R_F}{R_1} + \frac{R_F}{R_2}\right); \quad \beta = \frac{R_F}{R_1}; \quad \gamma = \frac{R_F}{R_2}.
\end{aligned}
$$

Copyright © 2005 IOP Publishing Ltd.

The circuit is a convenient way of **mixing** or adding signals. This feature is one reason making shunt feedback a popular configuration, used in a wide variety of applications.

Sketch V_{out} if V_1 is a DC voltage, V_2 an AC voltage and $V_3 = 0$.

An entertaining demonstration is to supply any two inputs from two sine wave generators very close in frequency; V_{out} exhibits beats because the phase difference of the two inputs varies with time. If you do not have two generators, it is easy to generate one of them by inductive pick-up of a 50 Hz signal from the mains; then tune your generator close to this frequency and adjust the two amplitudes to be similar in magnitude. Pick-up of mains signals is indeed all too common a problem.

7.6 The Differential Amplifier

A differential amplifier is the special case where $V_{out} \propto (V_3 - V_2)$. The input V_1 is eliminated and resistor values are adjusted so that $\alpha = \gamma$. This is achieved by setting $R_F/R_2 = R_4/R_3 = r$. The algebra simplifies to

$$V_{out} = r(V_3 - V_2). \tag{7.8}$$

This is a very valuable arrangement. Suppose V_3 and V_2 both contain an unwanted background of the same magnitude. It might be a mains signal or some similar interference. By taking the difference between V_2 and V_3, this background is removed. This trick is called **common mode rejection**. It is a feature built into many electronic instruments.

Two examples will illustrate its use. Firstly, suppose a weak radio signal (from a distant galaxy perhaps) is to be detected in the presence of local pick-up. One uses two identical detectors, both receiving the local pick-up but only one detecting the distant source. The differential amplifier eliminates the background and selects out the distant source. It improves the signal/background ratio.

Secondly, suppose one wants to detect the small electrical signal from a patients's heartbeat. The body acts as a good aerial and picks up a large background signal, as you can verify by connecting yourself across the leads of a sensitive oscilloscope. To get rid of this background, one rigs up a dummy aerial of resistors and capacitors, whose values are adjusted by trial and error to simulate the body. Then a differential amplifier will reject the background and select the small signal from the heartbeat.

All operational amplifiers are designed to have a high common mode rejection ratio ($CMRR$). This figure is the ratio G_- for $(V_3 - V_2)$ compared with G_+ for a signal $(V_3 + V_2)$ common to both inputs:

$$CMRR = G_-/G_+.$$

Copyright © 2005 IOP Publishing Ltd.

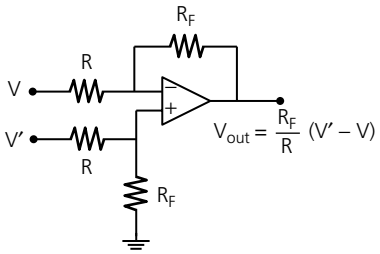


Fig. 7.10. Differential amplifier.

For the μA741C, it is 90 db or $2^{15} \simeq 33000$. What it impies is that the ideal performance of the operational amplifier expressed by equation (7.2) should be replaced by

$$V_{out} = G_-(V_+ - V_-) + G_+(V_+ + V_-) \qquad (7.2a)$$

with $G_+/G_- < 1/33000$.

In order to achieve the best common mode rejection, it is important that the input impedance should be the same for V_2 and V_3; otherwise, different proportions of these voltages will be dropped in the source resistances. The input impedance for V_3 is $(R_3 + R_4)$, since negligible current flows into the amplifier. What about V_2? The voltage at A on figure 7.9 is $V_3R_4/(R_3+R_4)$. So if $V_2 = V_3$, the voltage across R_2 is $V_3R_3/(R_3+R_4)$ and the input impedance for V_2 is $V_2/I_2 = R_2(R_3+R_4)/R_3$. For this to be equal to $(R_3 + R_4)$ requires that $R_2 = R_3$, hence $R_4 = R_F$. This arrangement is shown in figure 7.10.

It is instructive to make up the circuit of figure 7.9 and apply the same AC signals to V_2 and V_3. If the common mode rejection is perfect, there is zero output. Using variable resistors for R_F and R_2 and viewing V_{out} with the oscilloscope, you quickly get a feeling for common mode rejection.

7.7 Gain-Bandwidth Product

Shunt feedback gives an amplification

$$A \simeq -R_F/R_1 = -1/B.$$

As always, it is necessary to examine how good the approximations are. Obviously they must break down at some high frequency where $GB \to 1$. This condition gives the bandwidth, i.e. the range of frequencies over which the amplification is constant (within a factor $1/\sqrt{2}$). To demonstrate this result, it is necessary to go through the algebra keeping terms which involve G. We shall take the opportunity to examine at the same time the effect of the input impedance of the operational amplifier, r_{in}; it turns out to have negligible effect for normal values.

In figure 7.11 the operational amplifier is replaced by its Thevenin equivalent circuit, but omitting its output impedance. (If the latter is included, the algebra

Copyright © 2005 IOP Publishing Ltd.

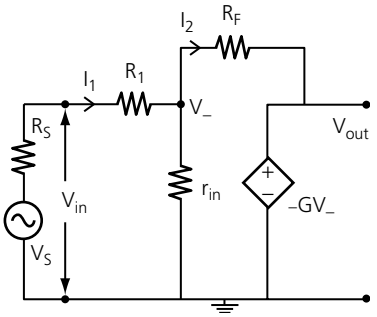


Fig. 7.11. Thevenin equivalent of Fig. 7.7.

gets rather messy, but exhibits no new features.) The following equations show a standard way of handling the algebra of shunt feedback. This procedure will be used with minor variants in subsequent examples. The voltage V_- at the negative input is expressed in as many ways as possible, four in this case:

$$V_- = V_S - I_1(R_1 + R_S) \tag{7.9}$$

$$= V_{out} + I_2 R_F \tag{7.10}$$

$$= (I_1 - I_2) r_{in} \tag{7.11}$$

$$= -V_{out}/G. \tag{7.12}$$

It is necessary to eliminate I_1 and I_2 and solve for V_{out}/V_{in}. From equations (7.9) and (7.12),

$$(R_1 + R_S) I_1 = V_S + V_{out}/G$$

and from equations (7.10) and (7.12),

$$R_F I_2 = -V_{out}(1 + 1/G).$$

Then equations (7.11) and (7.12) give

$$-\frac{V_{out}}{G} = \frac{r_{in}(V_S + V_{out}/G)}{(R_1 + R_S)} + \frac{r_{in} V_{out}}{R_F}\left(1 + \frac{1}{G}\right)$$

$$A = \frac{V_{out}}{V_S} = \frac{-R_F/(R_1 + R_S)}{1 + \frac{R_F}{G}\left(\frac{1}{R_F} + \frac{1}{(R_1 + R_S)} + \frac{1}{r_{in}}\right)}. \tag{7.13}$$

For large G and small R_S, the voltage gain is $-R_F/R_1$, reproducing the elementary result, equation (7.6). The bracket in the denominator is the admittance between the negative terminal of the amplifier and Earth. Writing $R_S + R_1 = R_1'$, the largest

Copyright © 2005 IOP Publishing Ltd.

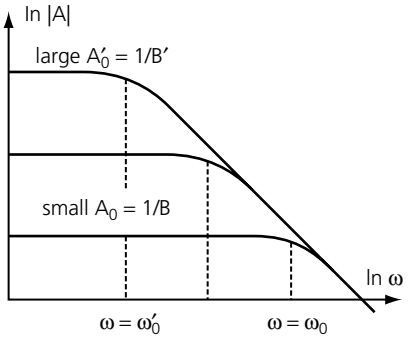


Fig. 7.12. Amplification v. ω.

correction term in the denominator is R_F/GR_1', which produces a correction to the voltage gain of $(1+R_F/GR_1')^{-1} = (1+1/GB)^{-1}$. If $G = 10^5$ and $B = 10^{-2}$, as in our example, figure 7.8(a), this correction term is 1 part in 10^3 at low frequencies. The remaining terms introduce even smaller corrections, since R_F and r_{in} are normally larger than R_1'. So if the dominant correction is retained, it is sufficient to take

$$A \simeq \frac{-R_F/R_1'}{1+R_F/GR_1'} = \frac{-1/B}{1+1/GB} = \frac{-1}{B+1/G} = \frac{-G}{1+BG}. \tag{7.14}$$

This is the same formula as for voltage feedback, equation (7.3a), except for a sign change. This sign change arises because the input is applied to the negative terminal of the operational amplifier; in the general formula of equation (7.3),

$$G_V = -G(\omega)$$

$$B = R_1'/R_F$$

and $$A = \frac{-G(\omega)}{1+BG(\omega)}.$$

Now consider the frequency dependence. At low frequencies, $BG \gg 1$ and the amplification is $A_0 = -1/B$ on the flat part of figure 7.12. This is called the **mid-band gain**: the name arises because at low frequencies the amplification falls again if AC coupling is used, as in figure 4.12. At high frequencies, $G(\omega)$ falls and eventually BG reaches 1. Around this point the amplification of the circuit begins to fall.

To examine this quantitatively, an expression is needed for $G(\omega)$. It is given by

$$G = \frac{G_0}{1+\mathrm{j}\omega/\omega_L}, \tag{7.15}$$

where $\omega_L \simeq 60$ rad / s for the μA741C. The factor j accounts for the *low* frequency cut-off; the operational amplifier itself is behaving as a low pass filter with lower

Copyright © 2005 IOP Publishing Ltd.

cut-off frequency ω_L. On the straight line part of figure 7.3(b), $G \simeq G_0/\mathrm{j}\omega$, where G_0 is a constant. So the bandwidth is given by

$$|BG| = |BG_0/\mathrm{j}\omega_0| = 1$$

$$\omega_0 = BG_0 = G_0/A_0.$$

The larger the value of the midband gain A_0, the smaller is the bandwidth ω_0. Indeed the simplest way of determining how G depends on ω is by measuring figure 7.12 experimentally by varying B. At frequency ω_0, the amplification drops by a factor $1/\sqrt{2}$ and a phase difference of 45° develops between input and output.

The product

$$\boxed{A_0\omega_0 = G_0} \tag{7.16}$$

is independent of B. It is called the **gain-bandwidth product** and is a fundamental parameter of the amplifier. A large bandwidth can be chosen or large amplification, but not both. Usually it is more difficult to achieve large bandwidth than large amplification. Often large bandwidth is important. In an oscilloscope, for example, it is vital that signals should be amplified by the same factor across the whole working range of frequencies. If necessary, large gain may be obtained by having more than one stage of amplification (within limits of noise and stability, to be discussed later). Thus it is commonplace to sacrifice amplification in order to achieve large bandwidth. These remarks apply equally well to voltage feedback, since the form of the voltage gain is precisely the same as equation (7.14), except for the sign.

The operational amplifier cannot provide any gain above the frequency where $G(\omega)$ falls to 1. This is called the **unity gain bandwidth**. For the $\mu A741C$ it is $\omega \simeq 6 \times 10^6$ rad/s. More expensive operational amplifiers have a unity gain bandwidth as high as $\omega = 5 \times 10^9$ rad/s.

7.8 Offset Voltage and Bias Current

So far the operational amplifier has been treated as ideal, except for the variation of G with frequency. Some practical limitations will now be discussed. Firstly, if there is no input to V_+ or V_-, the output voltage is not exactly zero; there is a small DC offset:

$$V_{out} = G(V_+ - V_- + V_{io}) \tag{7.17}$$

where V_{io} is called the **input offset voltage**. For the $\mu A741C$, $|V_{io}| < 1$ mV (and for some other operational amplifiers may be as low as 5 μV). Nonetheless, if $G = 10^5$, it may be sufficient to drive the amplifier to saturation when there is no input. An obvious question is what its effect will be including feedback.

Copyright © 2005 IOP Publishing Ltd.

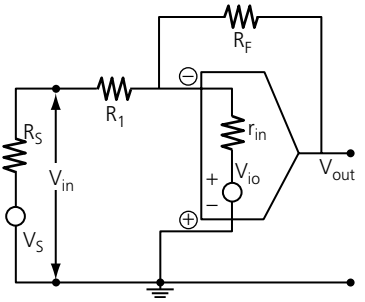


Fig. 7.13. Effect of V_{i0} with shunt feedback.

For series voltage feedback of Section 7.2, it is immediately obvious from equation (7.17) that the offset voltage simply adds to V_+, with the result that

$$V_{out} = \frac{(V_{in} + V_{io})G}{1 + GB}. \tag{7.18}$$

The effect is to introduce a DC shift into the output of $\sim V_{io}/B$; AC signals are superposed on this level. If $B > 0.01$, the DC shift is < 0.1 V and is unimportant; for larger amplifications it may become serious.

For shunt feedback, figure 7.13, the effect of V_{io} is to move the virtual earth at the negative terminal to voltage V_{io}. Then

$$\frac{V_{in} - V_{io}}{R_1} \simeq \frac{V_{io} - V_{out}}{R_F}$$

$$\text{so} \qquad V_{out} \simeq \frac{-R_F}{R_1} V_{in} + V_{io}\left(1 + \frac{R_F}{R_1}\right). \tag{7.19}$$

This introduces into the output a DC level of $V_{io}(1 + 1/B)$. If, for example, $V_{io} = 1$ mV and $B = 0.01$, the output DC shift is 100 mV.

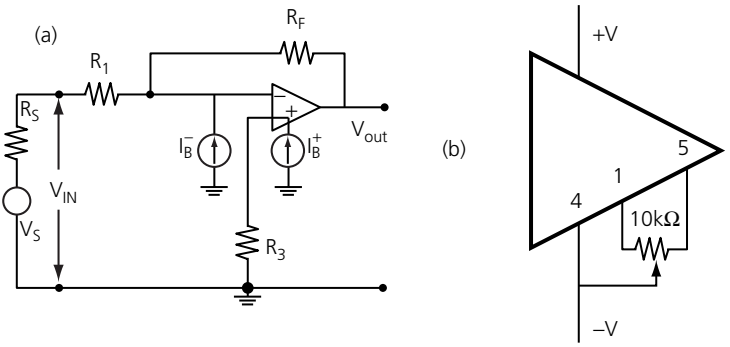


Fig. 7.14. (a) Input offset currents, (b) use of the offset null.

Copyright © 2005 IOP Publishing Ltd.

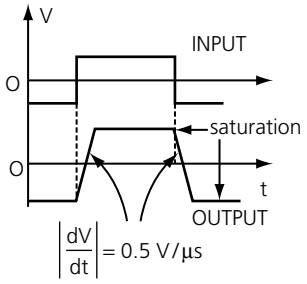


Fig. 7.15. Maximum rate of change of the output.

Bias currents

A second correction to ideal performance is that both input terminals of the operational amplifier draw small DC bias currents I_B^+ and I_B^-, indicated on figure 7.14(a) by the current generators. For the $\mu A741C$, $|I_B| < 200$ nA; but with $R_3 = 10^5\ \Omega$ this could simulate an offset voltage of 20 mV and produce a corresponding DC output shift. The moral is that large resistors should not be used directly in series with the input terminals of the operational amplifier. The manufacturer provides terminals 1 and 5, figure 7.15(b), where you can apply a correction via a resistor; this correction can be used to adjust the DC output to zero.

How can you measure (a) V_{i0} and (b) I_B^+?

Slew rate

A more significant limitation of the operational amplifier is that there is a maximum rate at which the output can change. This is called the **Slew Rate**. It is illustrated on figure 7.15. For the $\mu A741C$ it is 0.5 V/μs. If you feed in a sine wave, the output cannot rise faster than this and will be distorted at large amplitudes and large frequencies. It implies a limitation on the bandwidth for large signals. If $V_{out} = A \sin \omega t$, then $\mathrm{d}V_{out}/\mathrm{d}t = A\omega \cos \omega t$, and the slew rate S becomes a limitation when $A = S/\omega$. If A is equal to the full output range (15 V), the output of the μA741C is distorted above a frequency $\omega \simeq 3 \times 10^4$ rad/s called the **full power bandwidth**.

7.9 Complex Feedback Loops

Up to here the feedback has been via resistors. However, there is a multitude of useful circuits involving feedback via networks of resistors and capacitors. In figure 7.16, networks with impedances Z_1 and Z_F replace the resistors of figure 7.7. Then $V_{out}/V_{in} = -Z_F/Z_1$. This is called the **transfer function** of the circuit. Shaping the dependence of this transfer function on frequency can correct deficiencies in the frequency response of the source. This technique is called **frequency selective feedback**. We shall now discuss some examples.

Copyright © 2005 IOP Publishing Ltd.

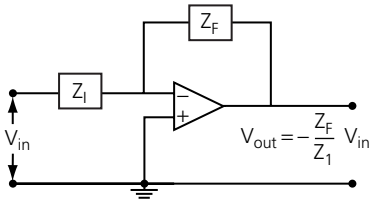


Fig. 7.16. Feedback using impedances Z_1 and Z_F.

Multipliers

Multipliers are available commercially. Four-quadrant multipliers accept two inputs of either polarity. They incorporate refined temperature stabilisation and accurate cancellation of input bias currents (which would upset the result).

Figure 7.17(a) illustrates how to do division and (b) shows how to form a square root. The multiplier is in the feedback loop and multiplies V_{out} and V_B in (a):

$$V_X = K V_{out} V_B = V_A;$$

$$\text{so} \qquad V_{out} = V_A / K V_B.$$

In (b),

$$V_X = V_A = K V_{out}^2$$

$$\text{and} \qquad V_{out} = \sqrt{V_A / K}.$$

Worked example

At this point, plenty of practice is desirable in handling voltage and shunt feedback, and your attention is directed towards problems 1–6 at the end of the chapter. Here, one worked example will show the way. The circuit of figure 7.18 will be analysed assuming the input is an AC signal of angular frequency ω. The approach is to

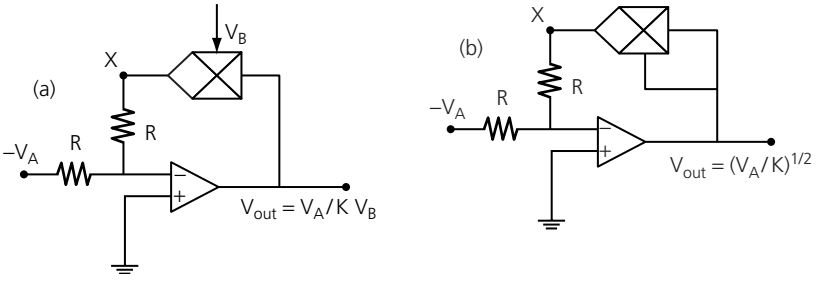


Fig. 7.17. (a) Using a multiplier to do division, (b) forming a square root.

Copyright © 2005 IOP Publishing Ltd.

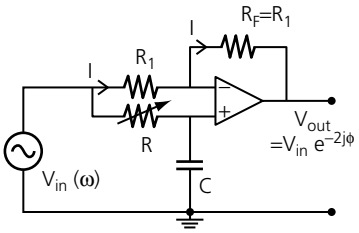


Fig. 7.18. Phase shifting ciruit.

make use of the approximation $V_+ = V_-$, so the first step is to find V_+. If the impedance $1/\mathrm{j}\omega C$ of the capacitor is written as Z_C,

$$V_+ = \frac{V_{in} Z_C}{Z_C + R} = \frac{V_{in}}{1 + R/Z_C} = \frac{V_{in}}{1 + \mathrm{j}\omega C R}.$$

Then

$$R_1 I = V_{in} - V_- \simeq V_{in} - V_+.$$

If the current flowing into the operational amplifier is negligible, the current through R_F is equal to I, so

$$V_{out} = V_+ - R_1 I = 2V_+ - V_{in}$$

$$= \frac{V_{in}}{1 + \mathrm{j}\omega C R}(2 - 1 - \mathrm{j}\omega C R) = \frac{V_{in}(1 - \mathrm{j}\omega C R)}{1 + \mathrm{j}\omega C R}$$

$$= \frac{V_{in}\{1 + \omega^2 C^2 R^2\}^{1/2} \mathrm{e}^{-\mathrm{j}\phi}}{\{1 + \omega^2 C^2 R^2\}^{1/2} \mathrm{e}^{\mathrm{j}\phi}} = V_{in} \mathrm{e}^{-2\mathrm{j}\phi}$$

where $\phi = \tan^{-1} \omega C R$. This circuit retards the phase of V_{in} by an angle 2ϕ without any change in magnitude.

This example is typical of a host of tricks which can be played using feedback networks and operational amplifiers. Exercise 2 gives many more. If you look in hobby magazines or manuals you will find scores of examples. Derivation of the transfer function always follows similar lines to the one here. The essential steps are: (i) setting $V_+ \simeq V_-$, (ii) neglecting the current into the operational amplifier.

7.10 Impedance Transformation

Figure 7.19 shows a circuit whose input impedance takes a curious and useful form, which can be used to manipulate impedances. The usual assumption will be made that the amplifier has large gain and high input impedance, so that the

Copyright © 2005 IOP Publishing Ltd.

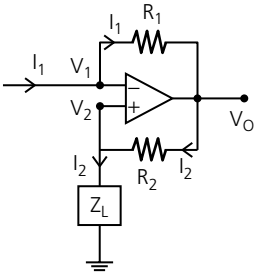


Fig. 7.19. An impedance transformer.

voltages V_1 and V_2 at the two input terminals are approximately the same. Then

$$I_1 = (V_1 - V_0)/R_1$$

$$I_2 = (V_0 - V_2)/R_2 = V_2/Z_L.$$

Setting $V_1 = V_2$ and dividing one equation by the other,

$$I_1 = -R_2 I_2/R_1 = -R_2 V_2/(R_1 Z_L) = -R_2 V_1/(R_1 Z_L);$$

the input impedance of the circuit is

$$R_{in} = V_1/I_1 = -R_1 Z_L/R_2. \tag{7.20}$$

The curious feature is that the input impedance includes a minus sign. If the load Z_L is resistive, the circuit behaves as a **negative resistor**. Instead of dissipating power, it can provide controlled power to the circuit to which V_1 is connected. This is a consequence of the positive feedback provided by R_2 and Z_L.

If Z_L is a capacitor with impedance $-\mathrm{j}/\omega C$, the circuit behaves like an inductor with impedance $\mathrm{j}\omega L$ where $L = (R_1/R_2)/\omega^2 C$. In integrated circuits, it is impossible to form any significant inductance within the silicon. At one particular frequency, the circuit of figure 7.19 can be used to simulate an inductance; however, it has a different frequency dependence. A clever and more elaborate circuit, called the **gyrator** (exercise 9, figure 7.40) reproduces the frequency dependence of an inductor. Although it looks complicated, it is readily fabricated inside an integrated circuit.

7.11 Input and Output Impedances with Feedback

It is now time to examine in greater detail the subtleties of voltage and shunt feedback, in particular input and output impedances. A general idea of their values is very helpful in planning how to link circuits together.

Figure 7.20(a) reproduces figure 7.6, showing in somewhat greater detail the circuit of figure 7.4(b) for voltage feedback. The operational amplifier itself is

Copyright © 2005 IOP Publishing Ltd.

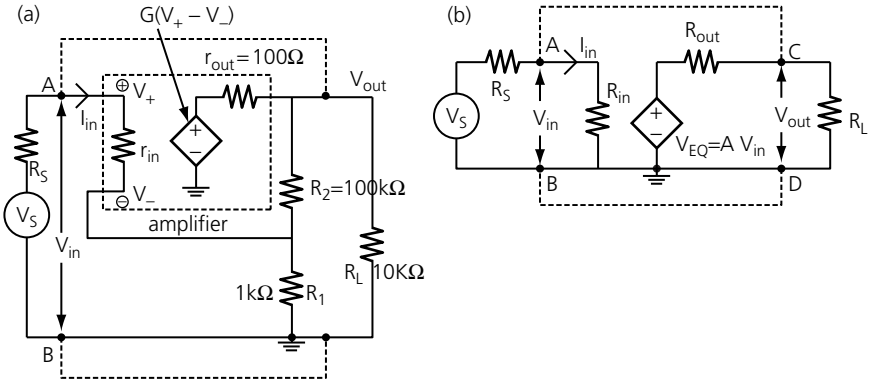


Fig. 7.20. (a) The circuit of the voltage feedback amplifier of Fig. 7.6 and (b) its equivalent circuit.

described as a 'black box' having input resistance r_{in} and an output drawn in Thevenin equivalent form with output impedance r_{out}. A load R_L is applied to the output and a source to the input.

For many purposes, it is useful to go further and represent everything inside the larger dashed rectangle by yet another black box, figure 7.20(b), having input impedance R_{in} and output impedance R_{out}. Once these two quantities are known, we can assess immediately the proportion of the source voltage V_S delivered to the circuit:

$$V_{in} = V_S R_{in}/(R_S + R_{in}) \tag{7.21}$$

and the fraction of the output voltage V_{EQ} delivered to the load resistor R_L:

$$V_{out} = \frac{A V_{in} R_L}{R_L + R_{out}}. \tag{7.22}$$

Input and output impedances R_{in} and R_{out} also play a second vital role. In association with series and parallel capacitances, they determine upper and lower cut-off frequencies, as discussed in Chapter 4. For example, R_{in} gives an idea of limitations to the bandwidth of the circuit from an educated guess about the capacitance across the input AB.

The input and output impedances R_{in} and R_{out} of the second black box are not the same as r_{in} and r_{out}, the values for the operational amplifier itself because of the feedback. Summarising what we shall find in the rest of the chapter, there are four important rules. The way to remember them is that they all make the DC performance of the amplifier better.

(a) Series voltage feedback increases the input impedance to

$$R_{in} = r_{in}(1 + GB);$$

Copyright © 2005 IOP Publishing Ltd.

if for example $GB = 10^5 \times 10^{-2} = 10^3$, this has a dramatic effect in increasing the input impedance. For a voltage amplifier, we want the input impedance to be large, so that most of the input signal appears across the input of the amplifier. Voltage feedback increases the input impedance and therefore helps in this respect. However, we shall also find that it reduces the bandwidth, which is less welcome.

(b) Shunt feedback decreases the input impedance; across the terminals of the operational amplifier itself,

$$R_{in} = r_{in}/(1 + GB).$$

This result, the **Miller effect**, again has dramatic consequences. For shunt feedback, the input signal is given by the current through the input resistor. Feedback increases R_{in} with the result that current diverted into the operational amplifier itself is reduced. Again this helps improve DC performance.

(c) If the feedback stabilises the output *voltage* by deriving the feedback from a voltage in the output circuit, the output impedance is reduced to

$$R_{out} = r_{out}/(1 + GB).$$

A low output impedance is desirable, so that little of the output signal is lost inside the amplifier. Again, feedback moves R_{out} in the right direction.

(d) If the feedback is derived from an output *current*, the output impedance is increased by a factor $1 + GB$. If one thinks of shunt feedback as operating on currents, one can think of the output as the Norton equivalent: a current source in parallel with R_{out}. Feedback increases R_{out}, so more of the output current is delivered to the following circuitry. Again this improves DC performance.

Input impedance

Let us begin with voltage feedback and figure 7.20. Consider R_{in} first. It is necessary first to solve for I_{in} hence $R_{in} = V_{in}/I_{in}$. With feedback, the signal across r_{in} becomes

$$V_{in} - V_{-} = V_{in} - BV_{out} = V_{in} - \frac{BGV_{in}}{1 + GB} = \frac{V_{in}}{1 + GB}. \tag{7.23}$$

The input impedance of the circuit is

$$\boxed{R_{in} = V_{in}/I_{in} = r_{in}(1 + GB).} \tag{7.24}$$

This demonstrates rule (a) given above.

Copyright © 2005 IOP Publishing Ltd.

Output impedance

The output impedance is deduced by applying Thevenin's theorem to the output CD of figure 7.20(b):

$$R_{out} = \frac{\text{open circuit voltage}}{\text{short circuit current}}.$$

The open circuit voltage is just $GV_{in}/(1 + GB)$ from equation (7.3a). When the output is short circuited, there is no feedback voltage, so $I_{out} = GV_{in}/r_{out}$. So

$$\boxed{R_{out} = \frac{GV_{in}}{1 + GB}\frac{r_{out}}{GV_{in}} = \frac{r_{out}}{1 + GB}.} \tag{7.25}$$

This result demonstrates rule (b).

These rules are valuable for low frequencies where G is large. However, at high frequencies where G falls to low values, the situation becomes messy and a computer simulation of the full equations is needed to give a complete understanding.

DC power supply

Figure 7.21 shows how to make a simple stabilised DC power supply using voltage feedback. The source V_S may be a battery or an unstabilised supply, and R_S can be large, so as to draw little current. The element D is a Zener diode, discussed in Chapter 1. It conducts in the opposite direction to the arrow when the voltage across it exceeds some well defined threshold value V_0, see figure 1.8(a). Providing $V_S > V_0$, a current flows through R_S and maintains a voltage V_0 at the input of the operational amplifier. This is amplified and $V_{out} = V_0(R_1 + R_2)/R_1$. The output voltage may be controlled by varying R_1 or R_2. The output impedance is low ($< 1\ \Omega$ if $GB > 100$). This is what is required of a power supply.

In practice, common operational amplifiers are limited in output current to 20–100 mA. If more current is required, a high power transistor must be added to boost the current. When operational amplifiers are used alone to drive a low impedance load such as a loudspeaker (standard values are 4, 8 and 12 Ω), the output current is likely to be limited by what the chip will supply. Most chips are protected internally against overload and will survive. However, a small AC signal will be amplified correctly, while a large one is limited by the chip. Large AC signals are then distorted.

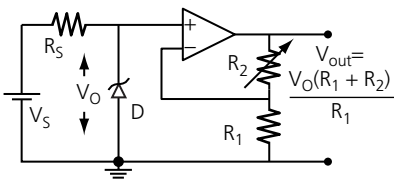


Fig. 7.21. A simple stabilised power supply.

Copyright © 2005 IOP Publishing Ltd.

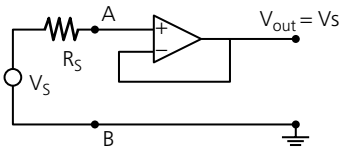


Fig. 7.22. A buffer amplifier.

The buffer amplifier

The high input impedance which results from voltage feedback means that the circuit draws very little current and acts as an almost ideal voltmeter. This may be a very desirable property. An extreme example of this is shown in figure 7.22, where *all* of the output voltage is fed back to the negative input terminal. In this case, $B = 1$ and the voltage gain is $A = 1$. This is called a **buffer amplifier** or **voltage follower** or **instrumental amplifier**. The input impedance of the circuit is $R_{in} \simeq Gr_{in} \simeq 10^5 \times 10^6\ \Omega = 10^{11}\ \Omega$. The output impedance is $R_{out} \simeq r_{out}/G \simeq 10^2/10^5\ \Omega = 10^{-3}\ \Omega$. It is possible, for example, to apply a standard cell to the input, giving $V_{in} = V_{st} = V_{out}$. Normally, no significant current can be drawn from a standard cell without polarising it. But the buffer amplifier is capable of supplying a large current at this reference voltage with negligible voltage loss in its output impedance. It is acting as a current amplifier. In chapter 11, the emitter follower will be discussed; this is a very common circuit which acts as a buffer amplifier in just the same way.

There is a snag associated with the large input impedance. It is impossible to avoid stray capacitances, both between the input terminals of the operational amplifier and between the positive terminal and earth. The very small current drawn from the source charges these capacitances only very slowly, so the bandwidth is poor. The associated bandwidth is $1/CR'$, where R' is the resistance appearing across the capacitance; from the Thevenin equivalent, R' is the parallel combination of R_{in} with R_S. If both R_S and R_{in} are high, the bandwidth is poor. For example, if $R_S = 10^9\ \Omega$ and $R_{in} = 10^{11}\ \Omega$ and $C = 50$ pF, the bandwidth is only 20 rad/s! This is a serious limitation. Shunt feedback greatly reduces the problem, and will be discussed shortly.

7.12 Stabilised Current Supplies

So far, feedback signals have been derived from the output *voltage*. One can arrange instead that the feedback signal is derived from the output current. If so, the output *current* is stabilised rather than the output voltage. This is the way to make a stabilised current supply, which might be used, for example, in supplying a magnet.

A simple way of doing this is shown in figure 7.23. In outline the circuit works as follows. Because $V_+ \simeq V_-$, the voltage across R_F is $\simeq V_{in}$ and current V_{in}/R_F

Copyright © 2005 IOP Publishing Ltd.

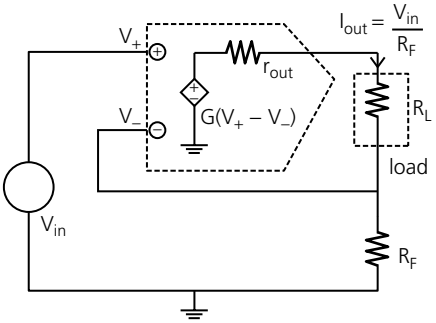


Fig. 7.23. Stabilisation of output current by voltage feedback.

flows through it. Very little of this current is supplied from V_- because of the high input impedance of the operational amplifier; almost all of it flows through R_L. The current through this load is stabilised at the value V_{in}/R_F, regardless of the magnitude of R_L.

7.13* Input Impedance with Shunt Feedback

The circuit for shunt feedback is drawn in greater detail in figure 7.24(a) showing both r_{in} and r_{out}. This whole circuit is now to be replaced with a Thevenin equivalent shown in figure 7.24(b).

In the absence of feedback, the input impedance is $R_1 + r_{in}$. With feedback, the negative terminal of the amplifier is a virtual earth, so the input impedance is approximately R_1. Feedback clearly reduces the effect of r_{in}. This is an important result worth remembering when linking circuits together.

The result may be derived algebraically from equations (7.9) to (7.12), but it is tedious and easy to make mistakes. It is more enlightening to follow a diagrammatic approach. In figure 7.25(a), the components of figure 7.24(a) are redrawn in a more symmetrical configuration. Consider the input impedance across the terminals AB. It is tempting to say that this will be R_1 in series with a parallel combination of r_{in} and R_F (if r_{out} can be neglected). This is not correct, because

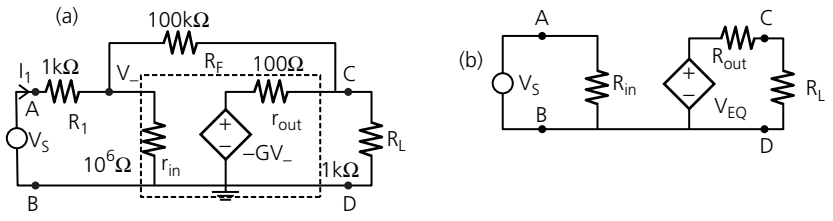


Fig. 7.24. (a) Amplifier with shunt feedback and (b) its equivalent circuit.

Copyright © 2005 IOP Publishing Ltd.

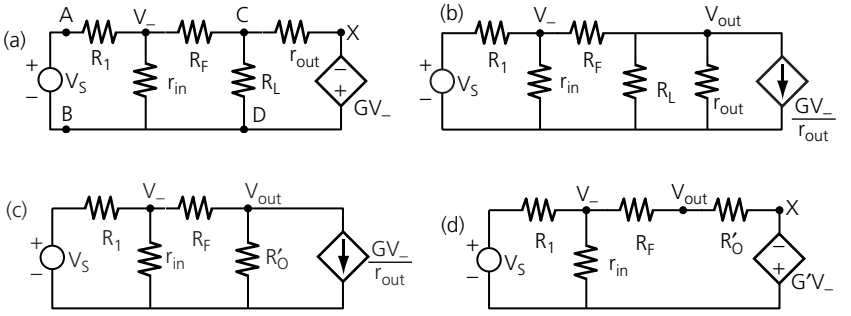


Fig. 7.25. Steps in simplifying the circuit.

the output generator GV_- produces a voltage which depends on V_- and Thevenin's theorem cannot be applied so simply. However, with some reorganisation of components, we can allow for this. In (b), the output generator and r_{out} are replaced by their Norton equivalent and in (c) the parallel resistors R_L and r_{out} are combined into $R_0' = r_{out}R_L/(r_{out} + R_L)$. Then (d) returns to the Thevenin equivalent with $G' = GR_0'/r_{out}$.

The point of these manoeuvres is that the voltage at X is known to be $-G'V_-$, so the voltage across $(R_F + R_0')$ is $(G' + 1)V_-$. Along this arm, earth potential is reached after resistance $(R_F + R_0')/(G' + 1)$. So from V_- to earth, the resistance is r_{in} in parallel with $(R_F + R_0')/(G' + 1)$, as shown on figure 7.26(a). This latter resistance is very small, of order 1 Ω if $R_F = 10^5$ Ω and $G = 10^5$; it bypasses r_{in} and is responsible for lowering the input impedance across the operational amplifier to a very low figure. Note, however, that G varies with frequency, and so does the resistance of the path bypassing r_{in}.

We need to distinguish carefully between the input impedance to the circuit across AB and the input impedance of the amplifier, across XB in figure 7.26(a). The former gives the fraction of the input signal delivered to the circuit by a source of resistance R_S, as in figure 7.26(b).

Suppose there is stray capacitance C across the input of the operational amplifier itself. The limitation from this on bandwidth is $\omega_0 = 1/CR'$, where R' is given by the parallel combination of R_1, r_{in} and $(R_F + R_0')/(G' + 1)$. Because the last of these is very small, there is hardly ever a real limitation compared with the

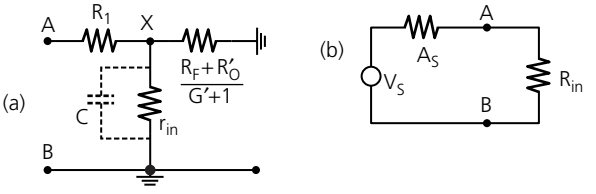


Fig. 7.26. (a) Input impedance with shunt feedback, (b) loading of the source.

Copyright © 2005 IOP Publishing Ltd.

natural bandwidth of the operational amplifier. In this respect, shunt feedback is superior to voltage (series) feedback, which increases the input impedance of the operational amplifier.

The algebra required to find the output impedance with shunt feedback is rather messy. The result is that the low output impedance of the operational amplifier (typically 75Ω) is reduced still further. For all practical purposes it is negligible. For completeness, the algebra is given in exercise 7.12.

7.14 Oscillation

The basic equation for negative feedback, equation (7.3a), states that the amplification with feedback is $A(\omega) = G(\omega)/(1 + BG(\omega))$. However, if it is possible for $G(\omega)B(\omega)$ to reach the value -1, $A(\omega) \to \infty$ and the circuit can produce an output without any external input. A familiar example is a public address system. Most of us will have witnessed occasions when the amplification is turned up and the system breaks into a high pitched scream. This is because the time delay between the sound going out from the loudspeakers and returning to the microphone produces a phase shift of 180°. So positive feedback arises at this frequency.

Care is needed to prevent something similar happening in electrical circuits. The voltage across a capacitor differs in phase by 90° from that across a series resistor. At high frequencies, there is inevitably stray capacitance across the output of the amplifier. If two amplifiers containing such stray capacitance are cascaded, a phase change approaching 180° can develop. This is not sufficient to cause oscillations because the signal across each capacitor goes to zero at high frequency. However, if three amplifiers are cascaded, a phase shift greater than 180° can arise. Any significant feedback from output to input can then lead to oscillation at the frequency where the phase shift is 180°.

Operational amplifiers are carefully designed to prevent this happening accidentally. In fact, the reason that operational amplifiers are designed with a gain falling linearly with frequency is to minimise the possibility of oscillation.

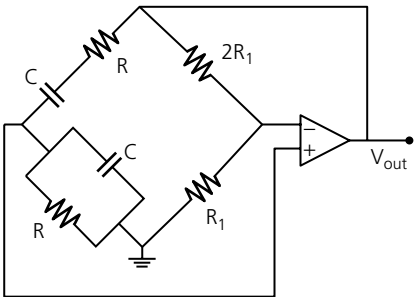


Fig. 7.27. Wien Bridge Oscillator, $f = 1/(2\pi CR)$.

Copyright © 2005 IOP Publishing Ltd.

If you wish to make an oscillator, at least two capacitors are required. An example is the Wien bridge oscillator of figure 7.27. The voltages at negative and positive inputs of the amplifier are

$$v_- = v_{out}/3$$

$$v_+ = v_{out} Z_2/(Z_1 + Z_2) = v_{out}/(1 + Z_1/Z_2)$$

where $$Z_2 = \frac{R/\mathrm{j}\omega C}{R + 1/\mathrm{j}\omega C} = \frac{R}{1 + \mathrm{j}\omega C R}$$

$$Z_1 = R + 1/\mathrm{j}\omega C = (1 + \mathrm{j}\omega C R)/\mathrm{j}\omega C.$$

So

$$v_+ = \frac{v_{out}}{1 + (1 + \mathrm{j}\omega C R)^2/\mathrm{j}\omega C R} = \frac{v_{out}\ \mathrm{j}\omega C R}{1 + 3\mathrm{j}\omega C R - \omega^2 C^2 R^2}. \tag{7.26}$$

The result is real if $\omega = 1/CR$. This condition defines the oscillation frequency. (The opamp needs to be chosen so that the slew rate is not a limitation at the operating frequency). At the oscillation frequency, $v_+ = v_{out}/3$. Oscillation requires $v_+ \geq v_-$, i.e. that the negative feedback is less than the positive feedback; this requires increasing the resistor $2R_1$ marginally to achieve oscillation. The output then saturates slightly at the peak and the output is not quite a perfect sine wave. In commercial generators, this problem is solved by stablising the output voltage or current below the saturation level.

7.15 Exercises

1. Find $A = V_{out}/V_{in}$ for the circuit of figure 7.28, assuming the amplifiers have high input impedance and low output impedance. (Ans: 65/12.)

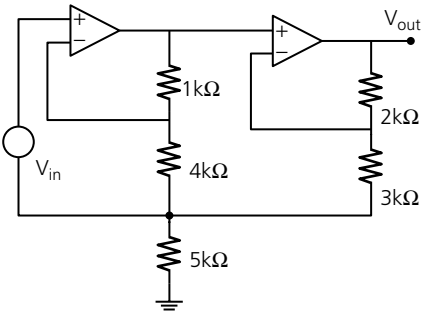


Fig. 7.28.

2. Demonstrate the relations given in figures 7.29–7.34. In every case, assume that the amplifier has a very high input resistance and a very high

Copyright © 2005 IOP Publishing Ltd.

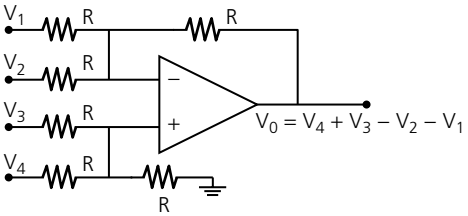


Fig. 7.29.

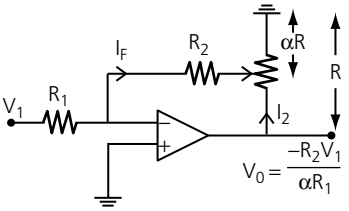


Fig. 7.30. Assume $I_2 >> I_F$

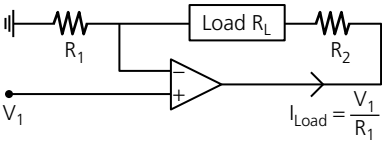


Fig. 7.31.

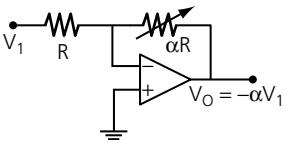


Fig. 7.32.

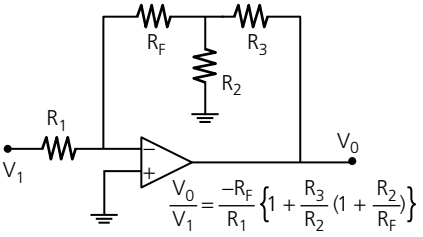


Fig. 7.33.

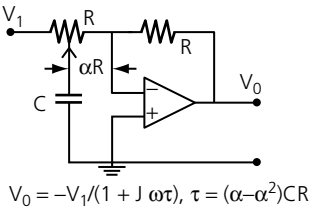


Fig. 7.34.

gain G, so that the voltage difference between the two inputs is very small. Note that the circuit of figure 7.33 can be used to generate a very large effective feedback resistance; this is useful if R_1 is large, and values of R_F required in figure 7.7 are too large to be readily available.

3. (King's). In the circuit of figure 7.35, the operational amplifier has a voltage gain A_1 of about -10^4, and A_2 is a difference amplifier whose output

Copyright © 2005 IOP Publishing Ltd.

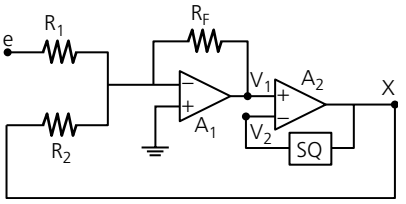


Fig. 7.35.

voltage is $A_2(V_1 - V_2)$ for inputs V_1 and V_2, where $A_2 \simeq +10^4$. SQ is a squaring circuit, whose output voltage is SV^2 for input V; $S \simeq +0.1$. Derive the relation which connects the output voltage x with the input voltage e. Assuming that $S = a$, $R_F/R_2 = b$ and $e = cR_1/R_F$, state the type of mathematical equation which this circuit solves and discuss any limitations which occur to you. (Ans: $ax^2 + bx + c = 0$; limitations: (i) requires $R_2/R_F << A_2$ and $A_1 >> (R_F/R_1 + R_F/R_2)$, (ii) it is not clear how to get the circuit to flip between the two solutions, and (iii) if $b^2 - 4ac < 0$, the roots of the equation are complex and the input $-(ax^2+bx+c)$ to A_2 is not zero, so A_2 saturates positively or negatively.)

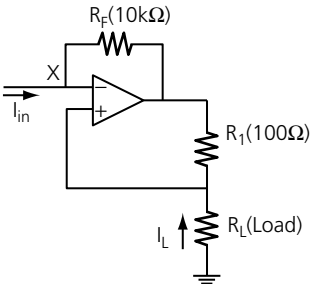


Fig. 7.36.

4* (IC). Show that the current I_L through the load resistor of figure 7.36 is proportional to the input current I_{in}, but independent of the value of R_L. The output current can be set using voltage V_{in} connected to point X via an input resistor R_{in}. Show that the input impedance of the circuit is $Z_{in} = V_{in}/I_{in} = R_{in} - R_F R_L/R_1$.

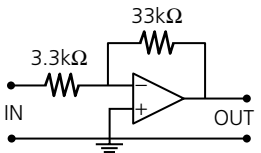


Fig. 7.37.

Copyright © 2005 IOP Publishing Ltd.

5. If in figure 7.37 the amplifier has a gain $G = 10$, an input impedance of only 1 kΩ and negligible output impedance, find the voltage gain of the circuit and its input impedance. How do these results change if $G = 100$? Hint: do this algebraically and substitute numbers at the end. (Ans: -1.85, $R_{in} = 4.05$ kΩ, $A = -6.94$, $R_{in} = 3.55$ kΩ.)

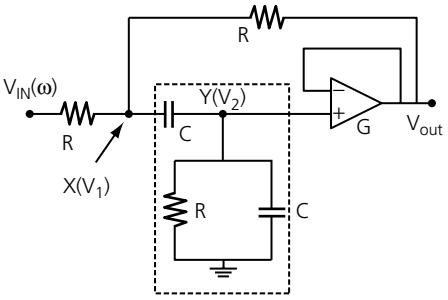


Fig. 7.38.

6*. (QMC). The amplifier in figure 7.38 has a large voltage gain G and very high input impedance. The input voltage at the left of the circuit is an AC signal of angular frequency ω. Show that the impedance Z of the network enclosed by the dashed lines is $Z = (2R + 1/j\omega C)/(1 + j\omega CR)$, and hence show that the voltages V_1 and V_2 at points X and Y are related by $V_2 = V_1/(2 + 1/j\omega CR)$. Show that $V_1(2 + R/Z) = V_{in} + V_{out}$. Hence, neglecting terms of order $1/G$, find V_{out}/V_{in}. Sketch the frequency dependence of V_{out}/V_{in}. (Ans: $V_{out}/V_{in} = 1/(4 + j\omega CR + 2/j\omega CR)$, figure 7.39.)

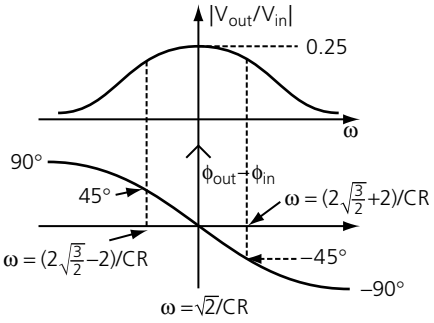


Fig. 7.39.

7. (QMC). What is meant by the bandwidth of an amplifier? An operational amplifier has high input impedance, a voltage gain of 10^5, and a bandwidth from angular frequency 0 to 10^3 rad/s (when used without feedback). How can this amplifier be used to provide (a) amplification

Copyright © 2005 IOP Publishing Ltd.

with a narrow bandwidth peaked at $\omega = 250$ rad/s? (b) amplification with bandwidth increased to cover the range $\omega = 0$ to 10^6 rad/s and an input impedance of 2.2 kΩ? In the latter case, what is then the voltage gain at low frequencies? (Ans: (a) tuned circuit as load with $LC = 1.6 \times 10^{-5}$. (b) Amplifier gain $\rightarrow$ 1 at 10^8 rad s^{-1}; use shunt feedback with $R_F/R_1 = 10^2$, $R_1 = 2.2$ kΩ, $R_F = 220$ kΩ, (c) 10^2.)

8. If in figure 7.13, $R_F = 330\,\text{k}\Omega$, $R_1 = 3.3\,\text{k}\Omega$, $r_{in} = 1\,\text{M}\Omega$ and $G = 10^5$, find the effect on V_{out} of an input offset voltage of 1 mV. If $I_B^+ = 20$ nA and $R_3 = 33$ kΩ in figure 7.15(a), what is the effect on V_{out}? (Ans: 100 mV, 0.66 V.)

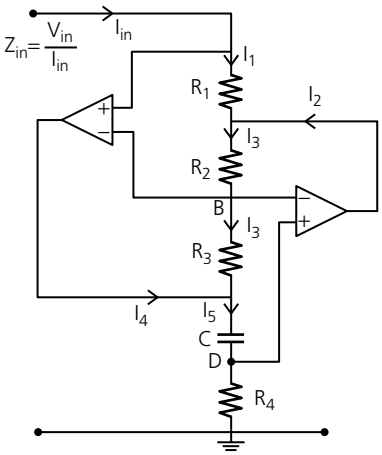


Fig. 7.40. The gyrator.

9. (RHBNC). (This problem is easier than it looks!) Hint: use the properties of operational amplifiers to relate the voltages at points B and D to the input voltage. Hence show that the circuit in figure 7.40 has an input impedance $Z_{in} = V_{in}/I_{in}$ given by $Z_{in} = j\omega C R_1 R_3 R_4/R_2$. Note that the gyrator behaves as an inductance $CR_1R_3R_4/R_2$. Inductors cannot be connected into micro-circuitry, so this is one way of simulating an inductor.

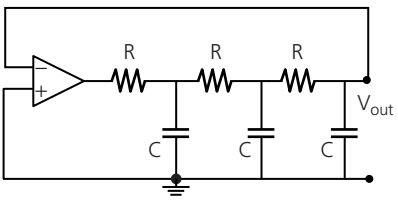


Fig. 7.41. Phase shift oscillator.

Copyright © 2005 IOP Publishing Ltd.

10. Show that for the phase shifting network of figure 7.41, $1/B = v_{in}/v_{out} = (1-5\omega^2C^2R^2)+\mathrm{j}(6\omega CR-\omega^3C^3R^3)$ and hence that for the oscillator of figure 7.41 the frequency of oscillation is $\omega = \sqrt{6}/CR$ and G_{mid} must be ≥ 29 for oscillation.

11*. A sensor detects 1 mV signals from a nerve cell. Its output impedance is 10^7 Ω. It feeds into an operational amplifier having gain 10^4, input impedance 10^5 Ω, output impedance 120 Ω, and unity gain bandwidth 5×10^6 rad/s. If voltage feedback is applied to the amplifier with $B = 0.01$, what is the magnitude of the output voltage? Remember that voltage feedback increases the input impedance of the operational amplifier by a factor $(1 + GB)$. What is the bandwidth if the input capacitance from the positive terminal of the operational amplifier to earth is 100 pF? What value of B is required for a bandwidth of 10^4 rad/s? What is then the output signal? (Ans: 50 mV, 2×10^3 rad/s; $B = 1/11$, 1.1 mV. The moral is that voltage feedback damages high frequency performance; shunt feedback is much better in this respect.)

12*. The objective of this exercise is to demonstrate that effects due to r_{out} in the feedback amplifier of figure 7.7 (shunt feedback) are extremely small. First show that $V_- = V'(R_F + R'_0)/(R'_1 + R_F + R'_0 + G'R'_1)$ where $V' = V_S R'_1/R_1$. Next, simplify figure 7.25(d) further by (i) redrawing V_S and R_1 in Norton equivalent form, (ii) combining R_1 and r_{in} in parallel to R'_1, and (iii) redrawing R'_1 and the current source in Thevenin equivalent form. Show that the current through the feedback resistor R_F is $I_F = (G' + 1)V_-/(R_F + R'_0)$. Hence show that the voltage gain is

$$\frac{V_{out}}{V_S} = \frac{-(R'_1/R_1)(R_F - R'_0/G')}{R'_1 + R_F + R'_0 + R'_1)/G'} \rightarrow \frac{R_F}{R_1} \text{ as } G' \rightarrow \infty.$$

Using $\mathrm{R}_f = 33$ kΩ, $\mathrm{R}_1 = 3.3$ kΩ, $\mathrm{R'}_{out} = 75\Omega$ and $G = 10^5$, what is the fractional effect due to (a) R'_0/G' in the numerator, (b) R'_0/G' in the denominator? (Ans: (a) 2×10^{-8}, (b) 2×10^{-7}).

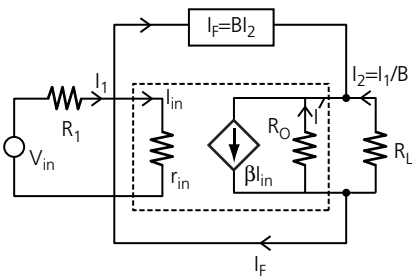


Fig. 7.42. A current amplifier.

Copyright © 2005 IOP Publishing Ltd.

13. Figure 7.42 shows a current amplifier. Demonstrate that

$$\frac{I_2}{I_1} = \frac{\beta}{1 + B(1 + \beta) + R_L/R_0} \tag{7.27}$$

and hence that $I_2 \simeq I_1/B$ for large β. From V_{in}/I_1, show that the input impedance to the circuit of figure 7.42 is $R_1 + r_{in}(B + 1 + R_L/R_0)/(B\beta + B + 1 + R_L/R_0)$. To find the output impedance of the circuit, consider terminals AB across R_L. Use equation (7.27) to find the voltage output across R_L with no additional load, i.e. V_{out} (open circuit); also use equation (7.27) to find the output current if R_L is short-circuited. From the ratio, show that the output impedance across R_L is R_L in parallel with $R_0(1 + B + B\beta)$.

Copyright © 2005 IOP Publishing Ltd.

8

Integration and Differentiation

8.1 Integration

Consider the circuit shown in figure 8.1. We shall find that, to a good approximation, it integrates the input voltage:

$$V_{out} \propto \int V_{in}\, \mathrm{d}t.$$

This is useful in many applications. If V_{in} is constant, it generates a linearly rising output which may be used to provide the timebase of an oscilloscope. A second example arises in the detection of ionising particles. Solid state detectors produce pulses of the form shown in figure 8.2 from the ionisation produced by a particle. The energy deposited is proportional to the area under the curve.

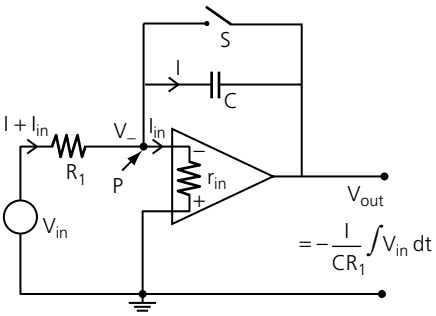


Fig. 8.1. The integrator.

To see roughly how the circuit works, suppose the operational amplifier has an input impedance r_{in} large compared with R_1 and a low output impedance. Then the current I_{in} into the amplifier itself is negligible and all of the current through R_1 flows into the feedback capacitor. Suppose that this capacitor has been discharged initially by closing and opening the switch S before the input pulse V_{in}

Copyright © 2005 IOP Publishing Ltd.

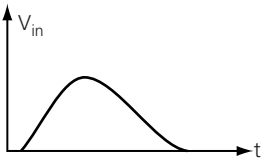


Fig. 8.2. Pulse from a solid state detector.

arrives. The point P at the negative input of the amplifier acts as a virtual earth and

$$I \simeq V_{in}/R_1$$

$$\boxed{V_{out} \simeq -\frac{Q}{C} = -\frac{1}{C}\int_0^t I\ \mathrm{d}t' = -\frac{1}{CR_1}\int_0^t V_{in}\ \mathrm{d}t'.} \tag{8.1}$$

The circuit is called an **integrator** and we refer to the **integration** of the input pulse. In a practical circuit the switch S is electronic and additional logic arranges to discharge the capacitor before use.

It is instructive to try this circuit experimentally. Suitable component values are $R_1 = 10\ \mathrm{k}\Omega$ and $C = 1\ \mu\mathrm{F}$. When V_{in} is zero, you will find that the output voltage drifts slowly positive or negative because of the input offset voltage or input bias currents. You can balance this drift with the arrangement described in the previous chapter in figure 7.14(b).

Suppose you now apply square pulses to the circuit. If their amplitude is perfectly symmetrical about zero, you should observe a triangular output $V_{out} = -(1/CR)\int V_{in}\ \mathrm{d}t$. Over the positive half of the input, V_{out} goes negative linearly with time, figure 8.3; during the negative half of the input, V_{out} should recover to zero with the opposite slope. In reality, the input is never accurately symmetrical about zero, so V_{out} drifts rapidly positive or negative and saturates. A diode placed across C will prevent this, but you may have to try both polarities to find the one which holds the DC level constant.

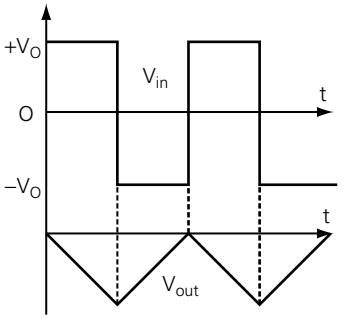


Fig. 8.3. V_{out} for an input bipolar square wave.

Copyright © 2005 IOP Publishing Ltd.

How is the diode functioning?

How good are the approximations in equation (8.1)? In fact, they turn out to be good if the duration of the input pulse is significantly less than the risetime of the circuit. The input impedance r_{in} of the amplifier will be included, but not its output impedance; the latter makes the algebra messy and adds no new features. We proceed in standard fashion, expressing the voltage V_- at point P in as many ways as possible:

$$V_- = V_{in} - (I + I_{in})R_1 \tag{8.2}$$

$$= I_{in} r_{in} \tag{8.3}$$

$$= V_{out} + (1/C) \int I \, \mathrm{d}t' \tag{8.4}$$

$$= -V_{out}/G. \tag{8.5}$$

In the last equation, the output impedance of the amplifier has been neglected.

In order to solve for V_{out}/V_{in}, it is necessary to eliminate I_{in} and I. From (8.3) and (8.5),

$$I_{in} = -V_{out}/G r_{in} \tag{8.6}$$

and from (8.2), (8.5) and (8.6),

$$I R_1 = V_{in} + \frac{V_{out}}{G}\left(1 + \frac{R_1}{r_{in}}\right).$$

Finally, from (8.4) and (8.5),

$$V_{out}(1 + \frac{1}{G}) = \frac{-1}{CR_1} \int \left[V_{in} + \frac{V_{out}}{G}\left(1 + \frac{R_1}{r_{in}}\right)\right] \mathrm{d}t'. \tag{8.7}$$

Corrections to the simple equation (8.1) are of order $1/G$. Faithful integration of a pulse therefore requires use of an *amplifier* with gain G extending to high frequency, so that the correction terms can be neglected.

Usually the input resistor R_1 is a few kΩ, and the input current is then a few mA. The bias current I_- at the negative terminal of the operational amplifier is typically 1 μA; for this reason, terms in equation (8.7) of the order $1/G$ are not trustworthy, particularly when you remember that G varies with frequency. Nonetheless, it is useful to follow through the algebra roughly what the effect of G is on performance.

Differentiating equation (8.7),

$$(G+1)\frac{\mathrm{d}V_{out}}{\mathrm{d}t} + \frac{1}{CR_1}\left(1 + \frac{R_1}{r_{in}}\right) V_{out} = -\frac{GV_{in}}{CR_1}. \tag{8.7a}$$

If V_{in} is a step rising from 0 at $t = 0$,

$$V_{out} = \frac{-GV_{in}}{1 + R_1/r_{in}}\left(1 - \mathrm{e}^{-\gamma t}\right) \simeq \frac{-tV_{in}}{CR_1}(1 - \tfrac{1}{2}\gamma t)$$

$$\gamma = \frac{1 + R_1/r_{in}}{CR_1(G+1)}. \tag{8.8}$$

Copyright © 2005 IOP Publishing Ltd.

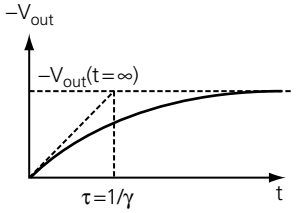


Fig. 8.4. $-V_{out}$ for constant V_{in}.

The shape of V_{out} against t is shown in figure 8.4, assuming no saturation. Deviations from a straight line occur for times of order $1/\gamma$. With $C = 1\ \mu$F, $R_1 = 10^3\ \Omega$ and $G = 10^5$, this time is 100 s. In reality, the output saturates long before this when V_{out} reaches the supply voltage. For the simple CR integrator of Chapter 3, γ was larger by a factor G. The amplifier has improved the linearity by the large factor G through the use of feedback. This is important in providing an accurately linear timebase for radar or an oscilloscope. The quantity γ is the reciprocal of the risetime of the output, and is equal to the bandwidth of the *circuit* (not the amplifier), as will be demonstrated below. Good linearity therefore requires small *circuit* bandwidth.

Are you clear about the distinction between circuit bandwidth and amplifier bandwidth? Remember the difference between the input impedance of the operational amplifier and from the point P of Fig. 8.5 to earth.

8.2 The Miller Effect

The integrator is a circuit where V_{out} depends on the past history of V_{in}, via the charge Q which has built up on the feedback capacitor. In a normal amplifier, this is undesirable; there the requirement is that V_{out} should be proportional to V_{in} itself, rather than its integral. Capacitance C_2 between input and output of the amplifier in figure 8.5 can seriously degrade the performance of the amplifier at high frequencies, where C_2 short circuits R_2. This is known as the **Miller effect**.

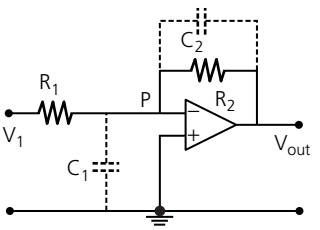


Fig. 8.5. Capacitance in a real amplifier.

Copyright © 2005 IOP Publishing Ltd.

To investigate it, suppose the input is an AC signal $V_1 e^{j\omega t}$. The output will be $V_0 e^{j\omega t}$, with V_0 complex. Ignore C_1 for the moment: it will become clear how to put it in later. If R_2 is absent, the relation between V_{out} and V_1 can be obtained from equation (8.7a):

$$V_0 = \frac{-GV_1}{1 + \frac{R_1}{r_{in}} + j\omega C_2(G+1)R_1} = \frac{-GV_1/R_1}{\frac{1}{R_1} + \frac{1}{r_{in}} + j\omega C_2(G+1)}. \tag{8.9}$$

The bandwidth ω_0 of the circuit is given by the value of ω for which

$$1/R_1 + 1/r_{in} = \omega_0 C_2(G+1)$$

$$\omega_0 = \frac{(1/R_1) + (1/r_{in})}{C_2(G+1)} = \gamma. \tag{8.10}$$

This verifies that the bandwidth of the integrator is γ.

The numerator of equation (8.10) can be interpreted in terms of a Thevenin equivalent. Viewed from the point P of figure 8.5, the resistance to earth (with R_2 absent) is R_1 in parallel with r_{in}. Call this R_1'. Then

$$\boxed{\omega_0 = 1/C_2 R_1'(G+1).}$$

The important point about this formula is that the circuit behaves as a filter where C_2 is scaled by the factor $(G+1)$. This is the Miller effect. The factor $(G+1)$ has the same origin as the factor $(G'+1)$ appearing in figure 7.26(a). Transistors often act as voltage amplifiers with $G > 100$. Their high frequency behaviour is then limited by capacitance from input to output. Therefore an essential feature of high performance transistors is low capacitance C_2 between input and output.

In figure 8.5, it is now clear how to handle the capacitance C_1; the input impedance of the operational amplifier becomes r_{in} in parallel with C_1. Finally, the effect of the feedback resistor R_2 may be included by replacing the admittance $j\omega C_2$ of C_2 with the admittance of C_2 in parallel with R_2. Then equation (8.9) becomes

$$V_0 = \frac{-GV_1/R_1}{\frac{1}{R_1'} + j\omega C_1 + \left(j\omega C_2 + \frac{1}{R_2}\right)(G+1)}.$$

At low frequencies, if the last term in the denominator dominates,

$$\frac{V_0}{V_1} \simeq \frac{-R_2}{R_1}\frac{G}{G+1} \simeq \frac{-R_2}{R_1}$$

which is the elementary result. The bandwidth ω_0 is given by

$$\omega_0\left(C_2 + \frac{C_1}{G+1}\right) = \frac{1}{R_2} + \frac{1}{R_1'(G+1)}. \tag{8.11}$$

Can you understand this result in terms of the parallel capacitances and resistances from the point P to earth?

Copyright © 2005 IOP Publishing Ltd.

For large G, $\omega_0 \simeq 1/C_2R_2$. So putting R_2 in parallel with C_2 (shunt feedback) has had the important and desirable effect of increasing the bandwidth by eliminating the Miller factor $(G + 1)$. Ultimately, G falls at high frequencies and C_1 and R_1' then come into play.

If you wish to demonstrate the Miller effect yourself on the oscilloscope, suitable component values are $R_2 = 1\ \text{k}\Omega$, $C = 0.01\ \mu\text{F}$.

8.3 Compensation

A common problem in pulse amplifiers is that stray capacitance distorts the leading edge of the pulse, either rounding it off or making it overshoot. Figure 8.6 shows possible responses to an input square pulse. The question arises how to cure this problem. In figure 8.6(d), a shunt feedback amplifier is shown with impedances Z_1, Z_2 and Z_{in} denoting parallel combinations of resistors and capacitors; e.g. Z_1 is C_1 in parallel with R_1. The gain of the circuit is given from the first form of equation (8.9) by

$$\frac{V_{out}}{V_{in}} = \frac{-G}{1 + (G+1)\frac{Z_1}{Z_2} + \frac{Z_1}{Z_{in}}} = \frac{-Z_2/Z_1}{1 + \frac{1}{G}\left(1 + \frac{Z_2}{Z_1} + \frac{Z_2}{Z_{in}}\right)}. \tag{8.12}$$

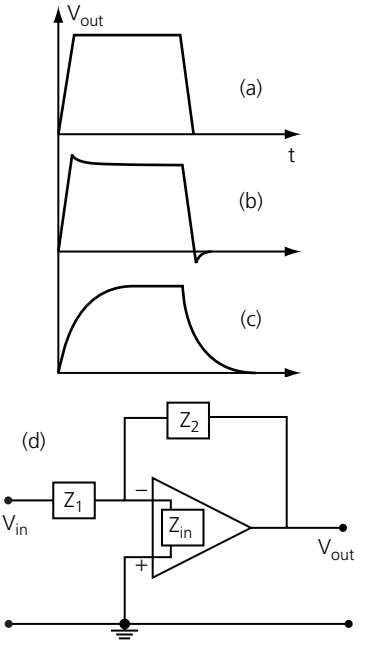


Fig. 8.6. V_{out} (a) with ideal adjustment, (b) non-ideal, overshooting, (c) non-ideal, reduced bandwidth, (d) compensation in the feedback amplifier.

Copyright © 2005 IOP Publishing Ltd.

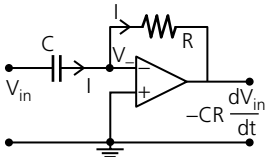


Fig. 8.7. Differentiating circuit.

If the gain is to be independent of frequency, it is necessary that Z_2/Z_1 and Z_2/Z_{in} should be independent of frequency. This requires that $C_1R_1 = C_2R_2 = C_{in}R_{in}$. Can you interpret this condition physically? In a practical situation, this condition can be achieved by adjusting small capacitors across two of the three locations until the best frequency response is obtained. If a square pulse is fed into the amplifier, these **trimming** capacitors are adjusted until the output shows the sharpest behaviour, figure 8.6(a), without overshooting, see exercise 5.

8.4 Differentiation

The circuit of figure 8.7 has the capacitor C and resistor R of figure 8.1 interchanged. It acts so as to differentiate the input waveform, as we now demonstrate. Making the usual assumption that the amplifier has large input impedance and small output impedance,

$$V_- = -V_{out}/G \simeq 0$$

$$V_{out} = V_- - IR \simeq -IR$$

$$V_{in} = V_- + (1/C)\int I\ \mathrm{d}t' \simeq (1/C)\int I\ \mathrm{d}t' = -(1/CR)\int V_{out}\ \mathrm{d}t'.$$

Differentiating this last equation

$$V_{out} \simeq -CR\mathrm{d}V_{in}/\mathrm{d}t. \tag{8.13}$$

There is, however, a problem with this arrangement. If there is noise at the input to the circuit or noise across any capacitance between the negative terminal and earth, the circuit differentiates it, producing a noisy output. Use of this circuit is not recommended. If you want to solve a differential equation by analogue techniques, it is best to express the equation in terms of integrating circuits.

Can you see physically why this circuit is noise sensitive?

8.5 The Charge Sensitive Amplifier

If the source driving an amplifier has large resistance, the circuit has a large time constant for charging any capacitance. We have already seen that this creates a

Copyright © 2005 IOP Publishing Ltd.

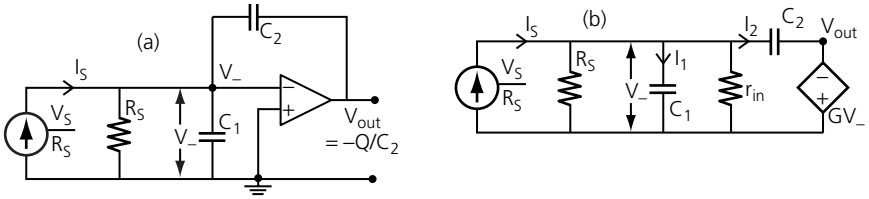


Fig. 8.8. (a) Charge-sensitive amplifier, (b) its equivalent circuit.

problem with bandwidth for series voltage feedback. The origin of the problem is that the input current is very small. A weak source may provide only a few electrons, so it is appropriate to think in terms of current and charge, rather than voltage.

Figure 8.8(a) shows a circuit which circumvents this problem. The source is expressed in Norton form, and (b) gives its equivalent circuit. If R_S is very large, all of the current $V_S/R_S = I_S$ flows into C_1 and C_2 (making the usual assumption that r_{in} of the operational amplifier is very large). Then

$$V_- = (1/C_1) \int I_1 \, \mathrm{d}t$$

$$V_-(1 + G) = (1/C_2) \int I_2 \, \mathrm{d}t$$

$$\int I_S \, \mathrm{d}t = \int (I_1 + I_2) \, \mathrm{d}t = V_-\{C_1 + C_2(1 + G)\}.$$

The output voltage is

$$V_{out} = -GV_- = \frac{-G}{C_1 + C_2(1 + G)} \int I_S \, \mathrm{d}t = \frac{-GQ}{C_1 + C_2(1 + G)} \qquad (8.14)$$

where Q is the charge generated by the source. For large G, $V_{out} \simeq -Q/C_2$.

This result is easy to understand. The impedance from the negative terminal to earth through C_2 is $1/\mathrm{j}\omega C_2(G + 1)$. The factor $(G + 1)$ makes this much smaller than the impedance of C_1, so nearly all of the current I_S flows to C_2 and creates a voltage $-Q/C_2$. The result is independent of C_1 so long as it is small compared with $C_2(1 + G)$. If the source is connected to the amplifier via a coaxial cable, V_{out} is insensitive to the capacitance C_1 of the cable.

Such an amplifier is called a **charge sensitive amplifier**. It is capable of amplifying very fast pulses, e.g. from a nuclear radiation detector, providing the transistors in the amplifier are fast enough. There is no bandwidth problem from capacitance. The only problem is that bias currents are liable to charge up C_2. It is necessary to cancel these out carefully and put a resistor across C_2, so as to discharge it with a time constant longer than the pulses of interest. This restores the DC level of the output after each pulse and prevents 'pile-up'.

Copyright © 2005 IOP Publishing Ltd.

8.6 Exercises

1. In figure 8.9, $R = 1$ kΩ, $C_2 = 100$ pF and $G = 200$. Find (a) the time constant of the output when V_{in} is a step function, (b) the bandwidth when V_{in} is an AC signal. If $C_1 = 10$ pF and a resistor $R_2 = 10$ kΩ is placed across C_2, what is the amplification of the circuit at low frequency? Use equation (8.11) to find the new bandwidth; notice that R_2 has had the effect of increasing the bandwidth greatly. (Ans: (a) 2×10^{-5} s, (b) 5×10^4 rad/s; $A = -10$; 10^6 rad/s.)

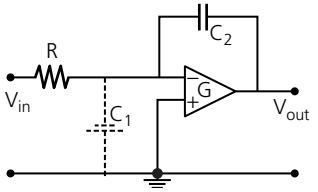
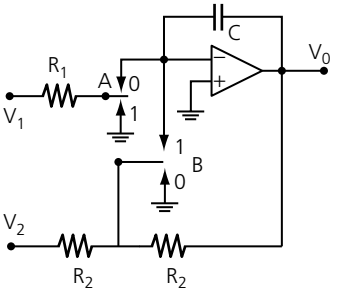


Fig. 8.9.

2. Demonstrate the relations given in figure 8.10, making the usual assumptions about the properties of operational amplifiers. In the reset stage, you should assume that the reset is applied for long enough for V_0 to settle to its final value. [In practice this last assumption will probably be upset by bias current I_1 in the operational amplifier.]



Operation	A	B
Reset	1	1
Compute	0	0
Hold	1	0

$$V_{out} = -\int_0^t \frac{V_1\,dt}{CR_1} - V_2$$

Fig. 8.10.

Copyright © 2005 IOP Publishing Ltd.

3**. (QMC). The amplifiers in figure 8.11 are DC wide-band inverting amplifiers. What output is produced by a 100 mV square pulse of duration 1 s? Hints:

i) Consider each half of the network separately. Assuming r_{in} of the operational amplifier is very high, show that

$$V_2\left(1+\frac{1}{G}\right) = -\frac{1}{CR}\int\left(V_{in}+\frac{V_2}{G}\right)\,\mathrm{d}t.$$

ii) Solve the problem for $t = 0$ to 1 s neglecting terms of order $1/G$.

iii) Solve the equation derived in (i) for V_2 exactly, with the trial substitution $V_2 = Ae^{\beta t} + B$, and show that the time constant is $CR(1+G)$.

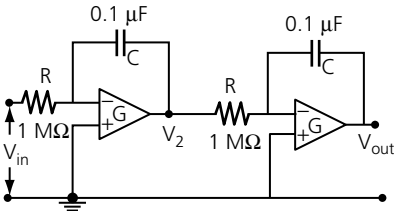


Fig. 8.11.

iv) Discuss in outline the output of the second stage for $t > 1$s. Will either amplifier saturate? If so, which one?

(Ans: (ii) $V_2 = -t$ V, $V_{out} = 5t^2$ V, (iii) $B = -GV_{in}$, $\beta = -1/CR(G+1)$, (iv) For $t > 1$, $V_2 \simeq -1$; V_{out} grows linearly to saturation.)

4. A 2-input analogue adder has a feedback resistance of 1 kΩ and operates from a ±15 V supply. Input through one resistor R_1 is 100 mV DC and through a second resistor R_2 is 300 mV RMS AC. If the output has an AC component 5 V RMS and a DC level -3 V, find R_1 and R_2. Sketch the output waveform if the input voltages are changed to 500 mV RMS AC and +400 mV DC. An integrator with a 33 nF capacitor is fed by a bipolar 100 mV peak-to-peak square wave input with 100 μs pulse width. If it is required to give an ouput triangular waveform of 3V peak-to-peak, what input resistor is required? (Ans: R_1 = 33Ω, R_2 = 60Ω; waveform of figure 8.12; 50 Ω.)

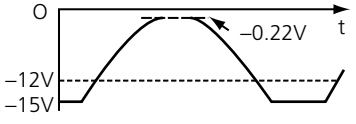


Fig. 8.12.

Copyright © 2005 IOP Publishing Ltd.

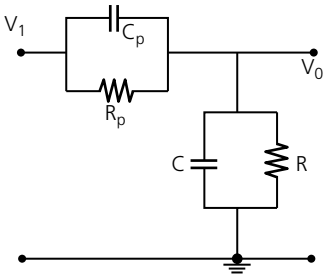


Fig. 8.13.

5. (QMC). Figure 8.13 shows the equivalent circuit of an oscilloscope probe. R is the resistance of the oscilloscope and C the total capacitance of the oscilloscope input and probe cable; V_1 is the probe input voltage and V_0 the voltage appearing on the oscilloscope. If $R_p = nR$ and $C_p = C/n$, show that (a) $V_0/V_1 = 1/(n+1)$ and is independent of frequency, (b) the impedance between the probe input terminals is equivalent to a resistance of $(n+1)R$ in parallel with a capacitance of $C/(n+1)$. If V_1 is a square pulse, figure 8.14 sketches the signal observed on the oscilloscope. Show that, after the capacitors have settled, $V_o = V_1/(1 + R_p/R_1)$. The height of the initial rise or fall of V_0 may be obtained from the high frequency impedance of the capacitors. Show that this is $V_1/(1 + C/C_p)$. Hence show that case (c) on the diagram corresponds to $C_p > C/n$, and case (d) to $C_p < C/n$.

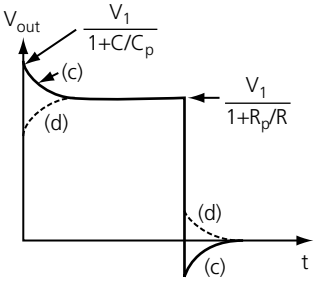


Fig. 8.14.

6. If in figure 8.6(d) Z_1 is a parallel combination of C_1 with R_1 and Z_2 and Z_{in} are parallel combinations of C_2 with R_2 and C_{in} with R_{in}, consider the effects of C_1, C_F and C_{in} on the phase of V_{out} with respect to V_{in}. Show that the primary effect of C_2 is to retard the output phase and of C_1 is to advance it; what is the effect of C_{in} on the output phase? (Ans: retards it.)

Copyright © 2005 IOP Publishing Ltd.

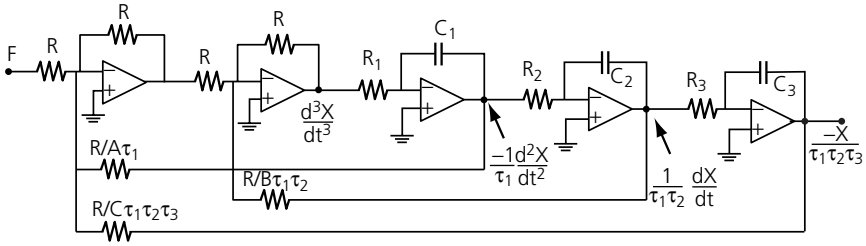


Fig. 8.15.

7. (Westfield). The circuit of figure 8.15 can be used as an analogue computer to solve differential equations (*if* bias currents in the amplifiers can be kept under control). Suppose the output of the second opamp is called $\mathrm{d}^3x/\mathrm{d}t^3$. Show that the outputs of succeeding opamps are as indicated by the arrows, where $\tau_1 = C_1R_1$, etc. Show that $\mathrm{d}^3x/\mathrm{d}t^3 = F(t) - Cx - B\mathrm{d}x/\mathrm{d}t - A\mathrm{d}^2x/\mathrm{d}t^2$. The output of the last opamp is then proportional to x.

Copyright © 2005 IOP Publishing Ltd.

9

The Diode and the Bipolar Transistor

9.1 Conductors

This chapter sketches the solid state physics involved in the operation of *pn* diodes and transistors. Just as it pays a driver to have an idea how the car works, so it is useful for the average physicist or engineer to have an outline understanding of how transistors work, plus an idea of typical performance figures. The physics is actually rather involved, so if you wish to grapple with the subtleties, it is necessary to refer to a specialised discussion, e.g. D.H. Navon, 'Semiconductor Microdevices and Materials', CBS Publishing Japan Ltd.

In a single atom, the Coulomb field of the nucleus produces a potential $V \propto Z/r$, where Z is the nuclear charge and r the distance from the nucleus. Electrons go into well defined energy levels, illustrated by the horizontal lines in figure 9.1. These electrons actually shield the nuclear charge to some degree and modify the Coulomb field from a $1/r$ dependence, but that is a detail.

In a solid, the situation is broadly similar, except that electrons close to one atom feel a potential due to neighbouring atoms as well. Low lying, tightly bound levels are affected very little, but the higher levels broaden into **bands**, illustrated in figure 9.2.

In a **conductor**, the least tightly bound electrons, called **valence** electrons, only partially fill a high lying band called the **conduction band** C on figure 9.2. This

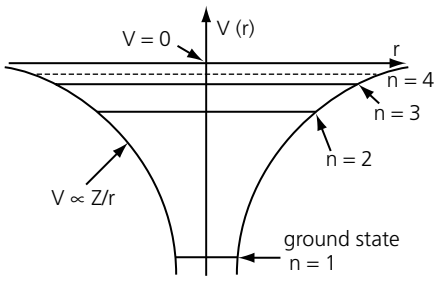


Fig. 9.1. Energy levels in an atom.

Copyright © 2005 IOP Publishing Ltd.

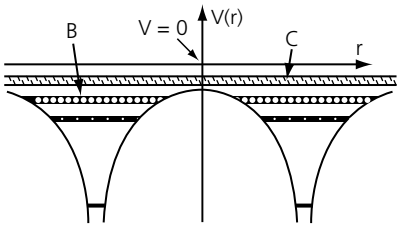


Fig. 9.2. Energy levels in a solid.

band is not confined to a single atom. Electrons in it are free to travel through the solid under the action of an electric field and hence carry current.

At absolute zero, the electrons occupy this band up to an energy E_F, called the **Fermi energy**; all states below E_F are occupied, all above E_F are empty. Above absolute zero, some electrons are excited by thermal energy into higher levels. The probability F_e that an energy level is occupied is shown in figure 9.3. Curves show the **Fermi-Dirac distribution**, derived from statistical mechanics; it is given by

$$\boxed{F_e(E) = \frac{1}{1 + \exp\{(E - E_F)/kT\}}} \tag{9.1}$$

where T is the absolute temperature and k is Bolzmann's constant. The value of E_F is such that $F_e = 0.5$ for $E = E_F$. At low temperatures, the curve rises steeply through E_F; as higher temperatures, more electrons are excited, and the curve gets steeper.

Electrons in levels well below the Fermi energy (by an amount $\gg kT$) move through the conductor without making collisions; such a collision would require transfer of energy and momentum from an atom (or another electron) sufficient to excite the electron to an unoccupied level. Thermal energy is rarely enough. However, electrons near the Fermi energy can transfer energy freely in a collision, sometimes gaining energy sometimes losing it. For electrons at the Fermi energy, the mean free path for collisions is $\lambda = 4 \times 10^{-6}$ cm in copper, i.e. hundreds of atomic diameters.

Consider next conduction under the action of an applied electric field **E**. Along a wire, the energy levels of all electrons at coordinate x are altered by energy $eV(x)$, where e is the charge of the electron and

$$V(x) - V(x = 0) = -\int_0^x \mathbf{E} dx.$$

The equilibrium Fermi energy slopes down the wire as shown in figure 9.4. An electron at $x = 0$ which happens to move to the right has extra energy above the Fermi level. In a collision, it is then more likely to lose energy than gain it, so on average it falls towards the Fermi energy and the atom with which it collides gains vibrational energy. This energy dissipation accounts for the heating of the wire due to resistivity.

Copyright © 2005 IOP Publishing Ltd.

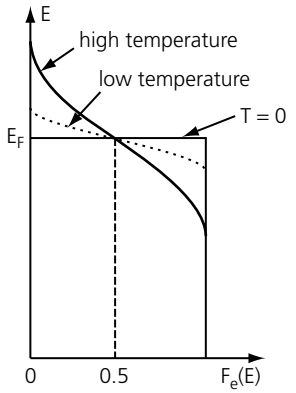


Fig. 9.3. The occupation probability F_e(E) of energy levels in the conduction band v. energy E.

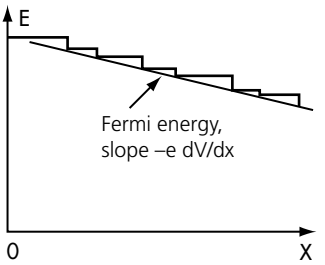


Fig. 9.4. Illustrating conduction.

9.2 Semiconductors and Doping

In an insulator, valence electrons fill a band, labelled V in figure 9.5; the energy gap E_G to the conduction band is so large (several eV) that there is a negligible number of excited electrons reaching unoccupied levels in the conduction band. This is why the material is an insulator. Semiconductors have a smaller energy gap (0.8–1.0) eV, and some electrons are excited thermally to the conduction band. However, the number is small enough that pure semiconductors still have a high resistivity.

When an electron is excited from the valence band it leaves a vacancy there. The probability distribution for vacancies is

$$F_v(E) = 1 - F_e(E) = \frac{\exp\{(E - E_F)/kT\}}{1 + \exp\{(E - E_F)/kT\}}$$

$$= \frac{1}{1 + \exp\{(E_F - E)/kT\}}. \tag{9.2}$$

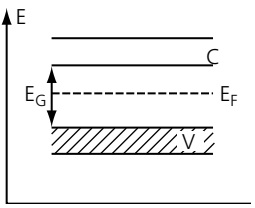


Fig. 9.5. The energy gap in an insulator or intrinsic semi-conductor.

Copyright © 2005 IOP Publishing Ltd.

This is the same as equation (9.1) except for the reversal in sign of $(E_F - E)$. On figure 9.5, the Fermi level lies midway between the top of the valence band and the bottom of the conduction band. This is because the few electrons excited into the conduction band leave a balancing distribution of vacancies in the valence band.

Doping

The resistivity of pure semiconductors is too high to be useful except in special applications. However, the conductivity can be raised to an interesting level by **doping**. This makes a semiconductor where the conductivity is due primarily to the impurities. Doping levels are typically 1 part in 10^6–10^8.

Suppose a crystal of silicon is doped with a small amount of a pentavalent material: phosphorus, arsenic or antimony. Silicon has a regular lattice in which each atom makes four covalent bonds to its neighbours. The impurity atoms are similar in size to silicon atoms, so they fit into the lattice by replacing them, as in figure 9.6. However, only four covalent bonds are made to neighbouring atoms. The fifth valence electron is very loosely bound by the unit charge of the impurity atom. In a hydrogen atom, the binding energy is 13.4 eV. But in silicon the dielectric constant is $\epsilon = 11.9$, so the binding energy is expected to be roughly $13.4/\epsilon^2 \simeq 0.1$ eV. The energy level is this amount E_N below the conduction band. This energy level is shown dashed on figure 9.7 to indicate that the electron is localised at the impurity atom. Experimentally, E_N varies from 0.039 to 0.049 eV.

At absolute zero, electrons fill energy levels up to and including the impurity level. The Fermi energy is roughly halfway between this and the conduction band. At room temperature, electrons are very easily ionised from the impurity level. Remember that at room temperature $kT \simeq 0.025$ eV, compared with $\frac{1}{2}E_N \simeq 0.02$ eV. Ionisation produces a fixed positively charged phosphorus atom and an electron in the conduction band, free to move through the crystal. The Fermi distribution of figure 9.7 has moved up close to the conduction band, implying much higher conductivity than for undoped material.

Semiconductors doped in this way are called **n-type** materials. The *n* signifies the fact that the impurities **donate** negative charges to the conduction band. Be

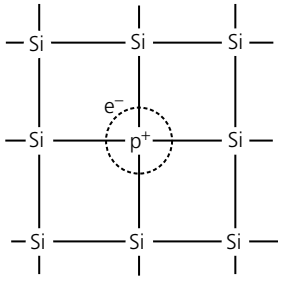


Fig. 9.6. n-type doping with phosphorus.

Copyright © 2005 IOP Publishing Ltd.

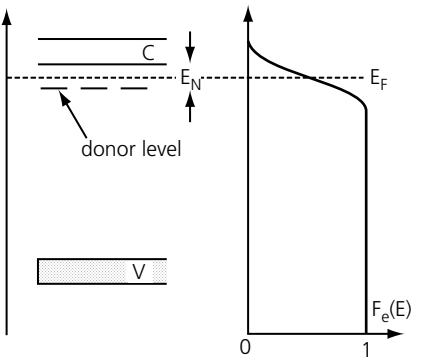


Fig. 9.7. Energy levels and $F_e(E)$ for an n-type material.

careful to distinguish this from the positive charge left on the lattice; the *n* refers to the mobile carrier.

p type material

The converse happens if silicon is doped with a trivalent material: boron, aluminium, gallium or indium. The impurity atom now steals an electron from a neighbouring atom, which in turn steals one from another atom and so on, figure 9.8(a). The result is what is called a **hole**: a missing electron. As the deficiency moves from atom to atom, the hole moves just like a positive charge. Such a material is referred to as **p-type**, because it generates positively charged carriers.

Draw a lattice of atoms with 4 electrons on all atoms except one, which has 3, figure 9.8(b). Throw dice to decide which way the hole moves. What happens when the hole reaches a boundary which is positively or negatively charged? What happens if two holes arrive at neighbouring sites?

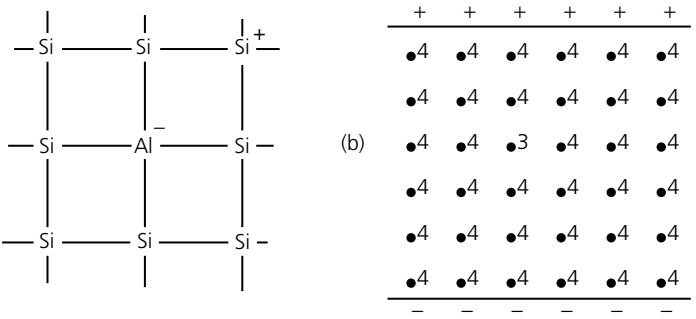


Fig. 9.8. (a) p-type doping. (b) a model of a lattice bounded by charged surfaces.

Copyright © 2005 IOP Publishing Ltd.

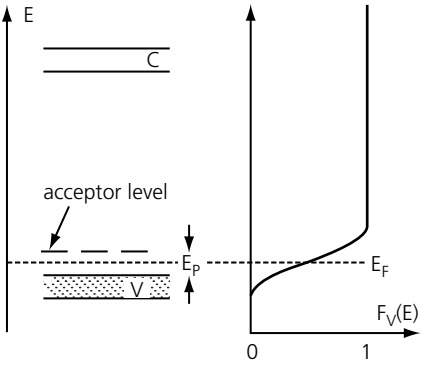


Fig. 9.9. Energy levels and $F_V(E)$ for a p-type material.

At absolute zero, electrons fill energy levels up to the top of the valence band, figure 9.9, and the impurity atoms, called **acceptors**, are unable to steal an electron. The impurity levels now sit empty ∼0.05 eV above the valence band. At non-zero temperatures, electrons are easily excited into them from the valence band leaving vacancies or holes. The Fermi energy is midway between the top of the valence band and the acceptor level; the probability distribution for vacancies, $F_v(E)$ of equation (9.3), is as shown in figure 9.9. If an electric field is applied, the hole moves just like a positive charge, but with a mobility which is a factor 3 smaller in silicon than for electrons.

In a pure semiconductor, the density of electrons and holes is the same. At room temperature it is about 10^{10} cm^{-3}. This is called an **intrinsic** semiconductor. On the other hand, in n-type material free electrons dominate overwhelmingly over free holes. There the electrons are **majority carriers** and the holes **minority carriers**. In p-type material, the roles are reversed. These materials, where densities of holes and electrons differ, are called **extrinsic** semiconductors.

9.3 The pn Junction Diode

A diode may be created from the junction between *p*-type and *n*-type materials. The current in the diode varies with voltage according to figure 9.10, repeated from Chapter 1. It rises exponentially with voltage V according to

$$I = I_0\{e^{eV/kT} - 1\}. \tag{9.3}$$

This formula arises from the Fermi-Dirac distribution, as we shall see shortly. In a bipolar transistor, the base current follows the same dependence on voltage V_{BE} between base and emitter.

Both *n*-type and *p*-type materials are made by heating pure silicon and exposing its surface to the impurity with which it is to be doped. The dopant may take the form of a hot gas or a liquid (indium melts at 156°C). At elevated temperatures,

Copyright © 2005 IOP Publishing Ltd.

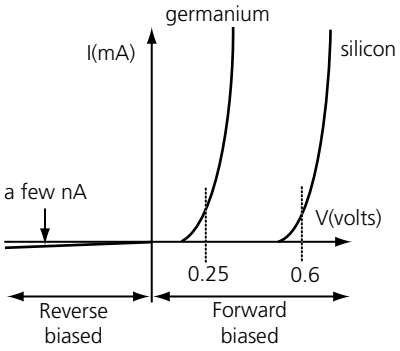


Fig. 9.10. Characteristics of pn diodes.

atoms diffuse through a solid in a fashion similar to a coloured dye diffusing through a liquid.

Suppose the silicon is initially exposed to donors, forming an n-type material. This is called the **substrate**. Then suppose the impurity is switched to acceptors. After a time these will outnumber donors near the surface of the silicon, converting it back to a p-type material, figure 9.11(a). This outlines how a pn junction is made. The physics of the interface between the two materials plays the decisive role. To simplify the discussion, it will be assumed that there is a sharp transition at the interface between p-type material, having a concentration of P acceptor atoms per unit volume to n-type material, with concentration N donor atoms per unit volume. It is not necessary that these doping concentrations should be the same. Of course, in a real material the junction is not sharp.

Diffusion of carriers

At room temperature, the impurity atoms themselves are immobile. However, electrons and holes diffuse readily across the junction. As shown in figure 9.11(b), electrons diffuse from the n-type region into the surface of the p-type material and recombine there with holes. Conversely, holes diffuse into the surface of the n-type material and recombine there with electrons. In the immediate vicinity of the junction there is a shortage of carriers of either type, because of recombination. This is called the **depletion layer**. In view of the shortage of carriers, it is a region of high resistivity and will support a large potential gradient.

Charge distribution

Diffusion of carriers leaves a surface layer of negative static charges in the p-type region, figure 9.11(c); these are the charges left on immobile acceptor atoms after holes have diffused away across the junction. Likewise, there is a surface layer of positive charge in the n-type region due to immobile donor atoms which have lost

Copyright © 2005 IOP Publishing Ltd.

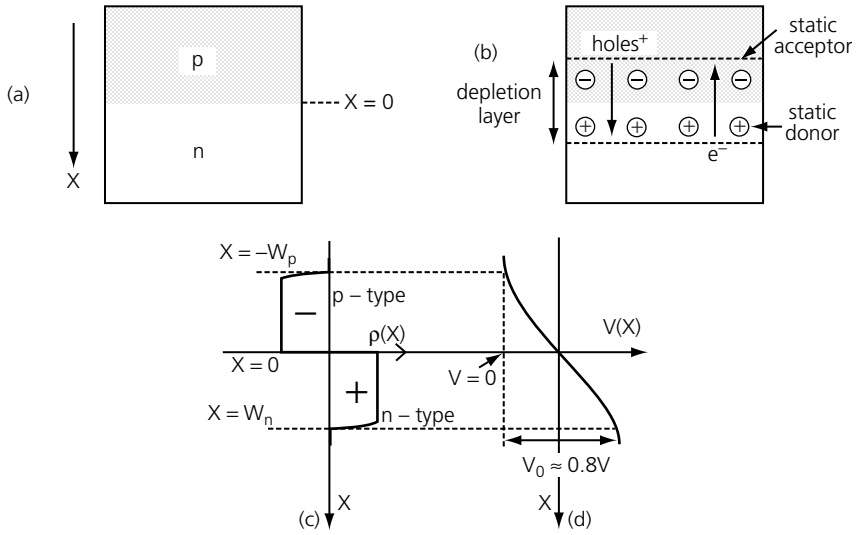


Fig. 9.11. (a) a pn junction, (b) the diffusion of mobile electrons and holes, (c) the charge distribution, (d) the potential V(x) across the junction.

electrons. Diffusion proceeds until these charges build up a potential difference V_0 across the junction sufficient to balance the diffusion. An electron which has diffused into the p region feels an electrostatic force pulling it back again. An equilibrium is established between diffusion one way and drift of carriers in the reverse direction due to the electric field created by the static charge distribution.

An electrostatic potential builds up across the junction. It is sketched in Fig. 9.11(d). It acts as a barrier for charges of either sign to climb. The charge distribution $p(x)$ of holes and the distibution $n(x)$ of electrons are

$$p(x) = P \exp\{-eV(x)/kT\} \tag{9.4}$$

$$n(x) = N \exp\{e[V(x) - V_0]/kT\}. \tag{9.5}$$

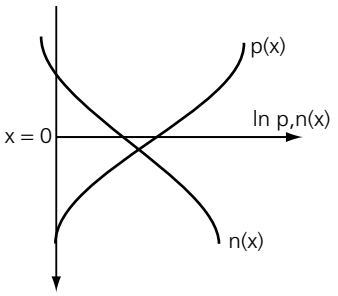


Fig. 9.12. p(x) and n(x).

Copyright © 2005 IOP Publishing Ltd.

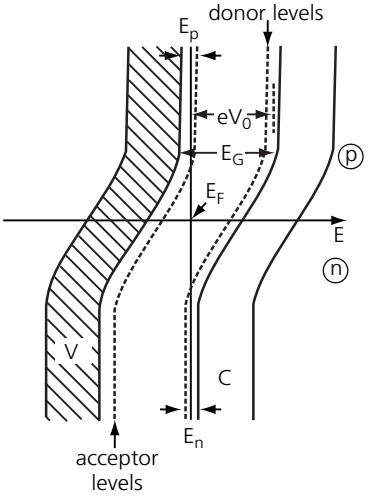


Fig. 9.13. Energy bands in the presence of V(x).

These densities vary rapidly with x and are shown on a log scale in figure 9.12. Both change by a factor $\exp\{-eV_0/kT\}$ over the width of the junction. If $V_0 = 0.8$ V, this is a factor 10^{14} at room temperature. A useful result from equations (9.7) and (9.8) is that the product $p(x)n(x)$ is independent of x:

$$p(x)n(x) \simeq PN \exp(-eV_0/kT). \tag{9.6}$$

This shows that either $p(x)$ or $n(x)$ must be small, and in the middle of the depletion layer both are small.

Energy levels across the junction are shown in figure 9.13. In equilibrium, the Fermi energy E_F must be flat and continuous across the junction. In the p region,

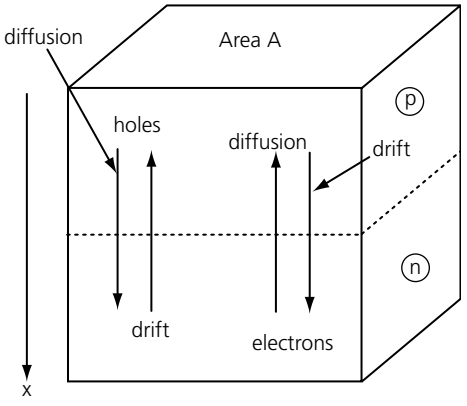


Fig. 9.14. Directions of diffusion and drift currents.

Copyright © 2005 IOP Publishing Ltd.

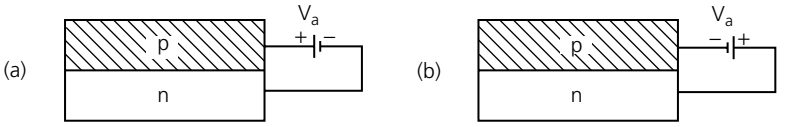


Fig. 9.15. (a) forward, (b) reverse bias.

E_F is close to the valence band V. In the n-region it is close to the conduction band C. The result is that $p(x)$ is very small in the n-region, because the probability of excitation from the valence band across the energy gap is small; likewise for $n(x)$ in the p-region.

Currents

At this point, figure 9.14 summarises the situation. There is an equilibrium where diffusion of holes and electrons is balanced by the static electric field across the material. What happens if this equilibrium is changed by applying an external voltage as in figure 9.15?

Effect of an applied voltage

Suppose a positive potential V_a is applied to the p side. This is called **forward bias**. It reduces the potential of figure 9.11(d) across the junction from V_0 to $V_0 - V_a$, figure 9.16. The density of electrons reaching the p side by diffusion rises from $N \exp\{-eV_0/kT\}$ to $N \exp\{-e(V_0 - V_a)/kT\}$, i.e. by a factor $\exp(eV_a/kT)$. A

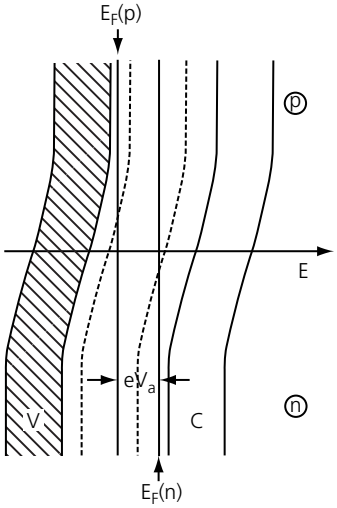


Fig. 9.16. Reduction of the potential across the junction by forward bias.

Copyright © 2005 IOP Publishing Ltd.

very few of these electrons actually make it to the terminal and flow round the external circuit. However, the vast majority neutralise in the p region with holes flowing in from the positive terminal of the battery; but the net result is still a current through the external circuit. The density of holes reaching the n side from the p-region likewise rises by the same factor. The total current across the junction is

$$\boxed{I = I_0\{\exp(eV_a/kT) - 1\}.} \tag{9.7}$$

Apart from the 1 on the right-hand side, ensuring that $I = 0$ when $V_a = 0$, I rises exponentially with V_a.

Diode characteristics

The resulting current is sketched in figure 9.17. It actually passes continuously through zero at $V_a = 0$ if plotted on a nA scale, as in (a). However, on a mA scale, the current appears to rise extremely rapidly around $V_a = 0.6$ V for silicon. For germanium, I_0 is larger because the energy gap is smaller; the rise on the mA scale for forward bias occurs for germanium at $V_a \simeq 0.25$ V.

Now consider reverse bias, figure 9.15(b). The potential across the junction is increased, figure 9.18. Because this potential acts as a barrier, the diffusion currents drop. If the bias is sizeable, the diffusion current drops essentially to zero and the reverse current is due to drift alone. On the p side of the junction there is a very small number of electrons according to equation (9.5). They see an electric field of the correct sign to accelerate them across the junction. Likewise, holes from the n side of the junction are accelerated across it. However, the drift current is very small because the concentration of these minority carriers is tiny. The reverse current is typically nA in silicon. There are useful side effects from this reverse current. These will be reviewed here before returning to diodes and transistors.

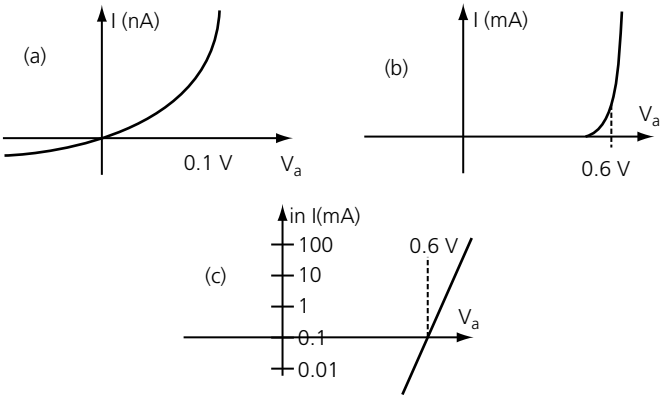


Fig. 9.17. Characteristics of a silicon pn diode, (a) on a nA scale, (b) on a mA scale, (c) plotted logarithmically.

Copyright © 2005 IOP Publishing Ltd.

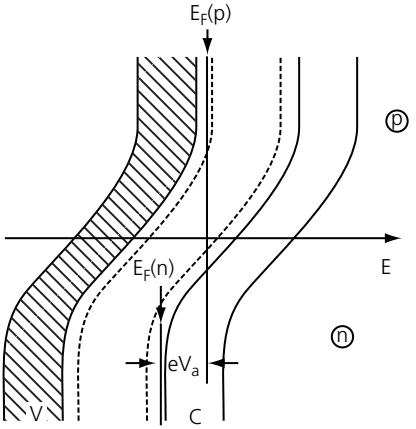


Fig. 9.18. Increase of the potential across the junction by reverse bias.

The PIN diode

Within the depletion layer, thermal processes spontaneously generate electron-hole pairs:

$$atom \rightarrow ion + electron \equiv hole + electron.$$

Both feel an electric field which sweeps them rapidly out of the depletion layer, so contributing to the reverse current. In the $p-i-n$ or PIN diode, the depletion layer is deliberately made of undoped intrinsic semiconductor. If light with an energy greater than the energy gap E_G is shone on to this depletion layer, figure 9.19, photons convert to electron-hole pairs and the reverse current measures the light intensity. This makes a **photodiode**. Its output impedance is high because almost all charges are swept out of the depletion layer, regardless of applied voltage.

Solid state counters

Alternatively, if an ionising particle traverses the depletion layer, it generates electron-hole pairs by ionisation. The carriers swept out of the depletion layer by the electric field create a pulse. The size of the pulse measures the energy deposit.

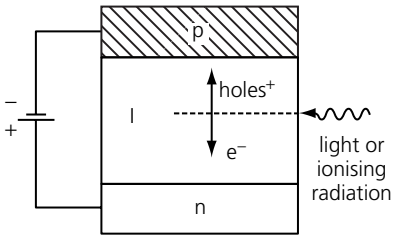


Fig. 9.19. Ionisation detector.

Copyright © 2005 IOP Publishing Ltd.

In Ge-Li and Si-Li semiconductor detectors, this is used to count particles and to measure energies with high resolution (1–2%). The pulse height is very insensitive to the magnitude of the reverse bias, which simply acts to sweep pairs out of the production region. Nearly pure silicon is used to minimise recombination. For Ge-Li and Si-Li detectors, the depletion layer can be several mm or even cm wide (at a cost!) if the applied voltage is high enough (several kV).

Zener diode

Two other processes affect the reverse current. If the mean free path of the carriers in the depletion layer is high, they may accelerate to sufficient energy to ionise atoms and generate an avalanche. Secondly, suppose the reverse bias is sufficient to distort the valence band of the p region level beyond the conduction band of the n region, figure 9.18. If the doping level is very high, the depletion layer is narrow enough ($<$100 Å) that electrons from the valence band on the p side can tunnel quantum-mechanically across the junction to the conduction band on the n side, creating a current. Both tunnelling and the avalanche process are used in the Zener diode to produce a sudden increase in reverse current above a well defined threshold voltage, figure 1.8.

9.4 The Diode as a Switch

We now review briefly some applications of the diode. Most depend on the fact that I is large for positive V and very small for negative V, i.e. the diode has a small forward resistance and large backward resistance. In this approximation it acts simply as a switch, allowing current to flow one way, but not the other. The idealised characteristic is then shown in figure 9.20. It requires 0.6 or 0.65 V to switch on a silicon diode.

Half-wave rectifier

An example of the use of a diode is in the half-wave rectifier shown in figure 9.21. Suppose the magnitude V_0 of the applied voltage $\gg$ 0.6 V. When the diode is

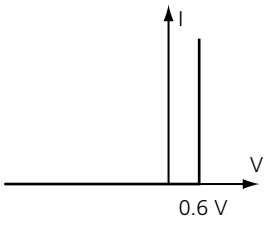


Fig. 9.20. Idealised diode characteristics (silicon).

Copyright © 2005 IOP Publishing Ltd.

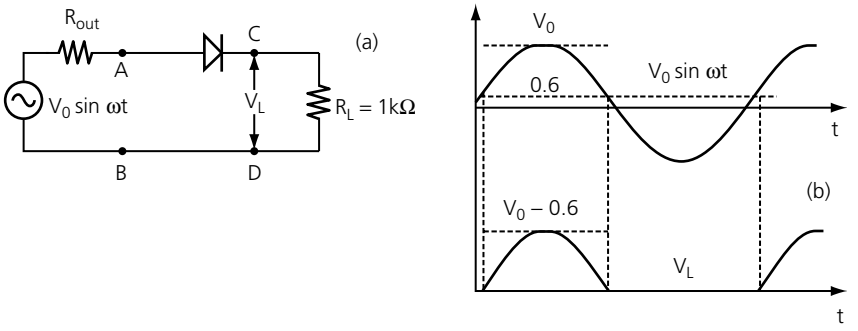


Fig. 9.21. (a) Half-wave rectifier circuit, (b) V_S and V_L v. t.

forward biased, the voltage across it is negligible, and almost all the voltage from the supply is applied to the load. On the other half-cycle, the diode is reverse biased, very little current flows, and all the supply voltage appears across the diode, very little across the load.

A **full wave rectifier** is shown in figure 9.22. The direction of current flow may be followed by drawing separate paths through the diodes for the two cases $V_S > 0$ and $V_S < 0$. The arrows in figure 9.22(a) indicate the current flow when $V_A > V_B$, and the waveforms are shown in (b). The diodes used in ordinary electronics carry currents of up to 10 or 100 mA. However, large diodes capable of carrying tens of thousand of amps can be made and are used in solid-state rectifiers for DC generators.

Power supplies

In a **DC power supply** it is necessary to smooth the ripple from figure 9.22(a) using a filter. The upper waveform of Fig. 9.22(a) may be smoothed as in figure 9.23 by using an inductor and large capacitor to make a low pass filter. The inductor presents zero impedance to DC but impedance $j\omega L$ to AC components. A DC voltage builds up on the capacitor.

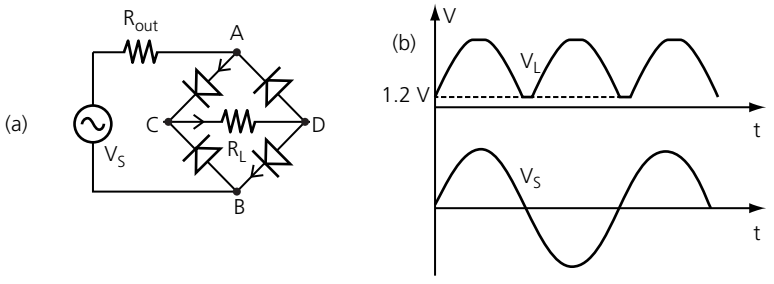


Fig. 9.22. (a) Full-wave rectifier, (b) waveforms.

Copyright © 2005 IOP Publishing Ltd.

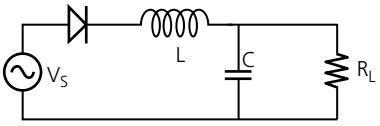


Fig. 9.23. Further filtering with an inductor and capacitor.

The output of these circuits depends on the magnitude V_0 of the supply voltage. If a **regulated** voltage supply is required, where the voltage is independent of V_0, this may be achieved as in figure 9.24 using a Zener diode, which breaks down at an accurately defined voltage V_1 and holds the voltage across R_L at this value. The difference between V_0 and V_1 is dropped across a large resistor R_1. Zener diodes may also be used to protect the input of a circuit against applied signals which are too large. For voltages below the threshold, no current flows through the diodes and the circuit behaves as if they were not there; above the threshold they conduct and bypass the rest of the circuit.

9.5 The npn Bipolar Transistor

The physics of a bipolar transistor is closely related to that of the diode; this is why so much attention was given to the physics of the diode. Figure 9.25 shows schematically the layout of the *npn* bipolar transistor. It consists of a *p*-type **base** and *n*-type regions called **emitter** and **collector**. The gap between the emitter and collector is made very small ($\sim 1\ \mu$m).

Typical characteristic curves are shown in figure 9.26. Here V_{CE} is the voltage between collector and emitter and V_{BE} the voltage from base to emitter. The device is controlled by V_{BE}. The base-emitter junction acts like a diode, so when the transistor is 'on' and current flows, the voltage at B is typically 0.6 V above that at E. The characteristic diode curve appears in figure 9.26(b). However, little current flows to or from the base. Most ($\geq$ 99%) flows between the emitter and collector, which sucks up all the electrons injected into the base from the emitter. The base acts as a control terminal which governs the emitter current. Indeed

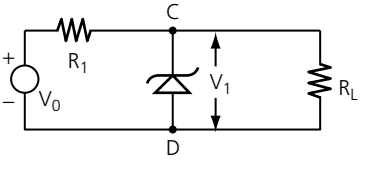


Fig. 9.24. Zener diode voltage regulator.

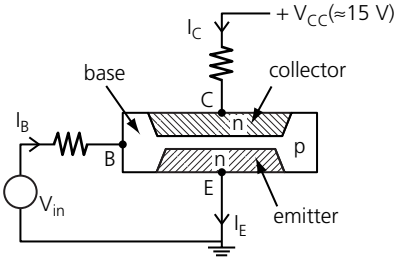


Fig. 9.25. The npn bipolar transistor.

Copyright © 2005 IOP Publishing Ltd.

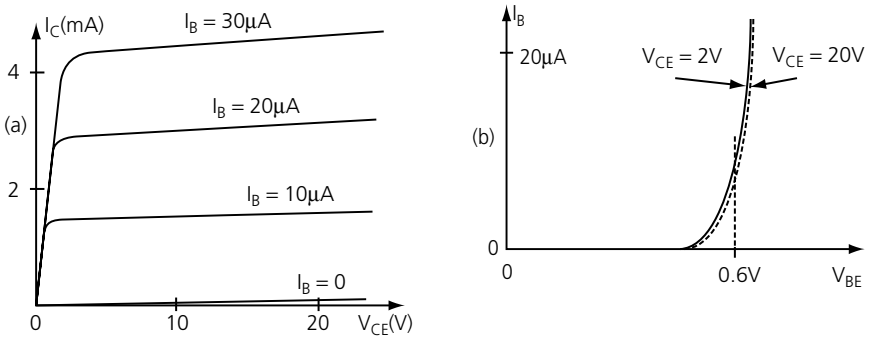


Fig. 9.26. Characteristics of the bipolar transistor.

the ratio β of collector current to base current is approximately constant over the normal operating range. Both sets of characteristic curves are insensitive to V_{CE}, providing it is positive and above about 0.25 V.

The emitter is heavily doped to ensure that most of the current is due to electrons diffusing from emitter (n-type) to base (p-type) rather than in the reverse direction.

Characteristic curves

If you wish to measure the characteristics of a bipolar transistor yourself, this is readily done with the circuit of figure 9.27. The symbol for the *npn* transistor is indicated in the middle of the figure, where collector, base and emitter are labelled C, B and E. The arrow indicates the controlling diode action between base and emitter. The transistor is said to be in the **common emitter configuration**, because the emitter is common to the input V_{BE} and output V_{CE}. The currents may be measured with AVOs or multimeters and the voltages with an oscilloscope or multimeter. The behaviour is very sensitive to V_{BE} and the current into the base should not exceed 40 or 50 μA for most transistors, so the 100 kΩ resistor is

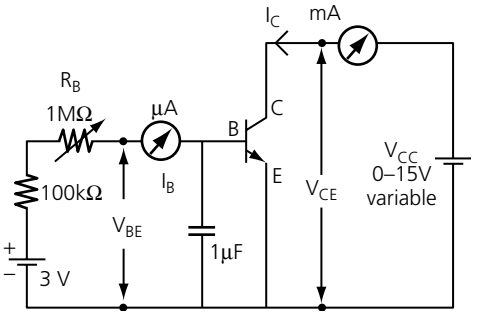


Fig. 9.27. Measuring transistor characteristics.

Copyright © 2005 IOP Publishing Ltd.

included for safety. A change in R_B alters I_B and the meter reading I_C shows a correspondingly much larger change in I_C. This demonstrates immediately that the bipolar transistor acts as a current amplifier. A practical point is that stray feedback from output to input can make the circuit oscillate at high currents; this is eliminated by the 1 μF capacitor which acts as a low impedance for high frequency signals.

In **summary**, the current delivered by the emitter is controlled very delicately by the voltage V_{BE} between base and emitter. However, because I_C/I_B is approximately constant, it is better to think of this type of transistor as a current amplifier. A small change in I_B goes with a tiny change in V_{BE} and a large change in I_C. The power in the collector circuit $V_{CE}I_C$ is very much larger than $V_{BE}I_B$; typically they are in the ratio 1000 to 3000. The essential virtue of the transistor is this power amplification: a small power in the base circuit controls a large power in the collector circuit.

As an aside, you might wonder whether the collector and emitter are interchangeable. In fact, they are to a limited extent. If you mistakenly invert their connections, the collector-base junction now acts as a diode and the transistor will function, but in an inferior way; this is because the collector is lightly doped compared with both base and emitter.

9.6 Simple Transistor Circuits

Emitter follower

Figure 9.28 shows a circuit called the emitter follower; it acts as a current amplifier, but gives no voltage amplification. If V_{in} is zero or negative, the diode is off and no current flows through the transistor, so $V_Y = 0$. If V_{in} is raised above 0.6 V the transistor switches on. Because the base-emitter diode conducts, $V_{BE} \simeq 0.6$ V and $V_Y \simeq V_{in} - 0.6$ V. This voltage appears across R_1 so

$$I_E = I_C + I_B = V_Y/R_1 = (V_{in} - 0.6)/R_1.$$

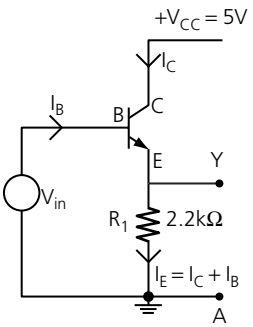


Fig. 9.28. A simple current amplifier.

Copyright © 2005 IOP Publishing Ltd.

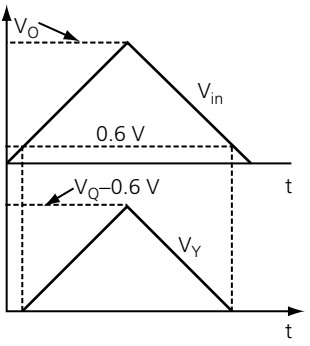


Fig. 9.29. Response to a triangular input voltage.

Figure 9.29 shows the resulting waveforms for a triangular input; V_Y follows V_{in} with a difference of 0.6 V. This is called **emitter follower** action. In this circuit there is no voltage amplification, but there is *current* amplification because $I_C = \beta I_B$, hence power amplification.

If you wanted the emitter follower to transmit both positive and negative halves of an AC input signal 2 sin ωt V*, what change to the circuit would achieve this?*

OR gate

A circuit very similar to figure 9.28 is used in digital logic. It is shown in figure 9.30. Two transistors share a common load resistor R_1. Inputs A and B are applied to the bases of the two transistors. If an input signal $>$ 0.6 V is present on either input, the point Y will follow it with a drop of 0.6 V between base and emitter. If inputs are present at both A and B, the point Y follows the larger input. This makes a rudimentary OR gate: there is an output present if there is a positive input

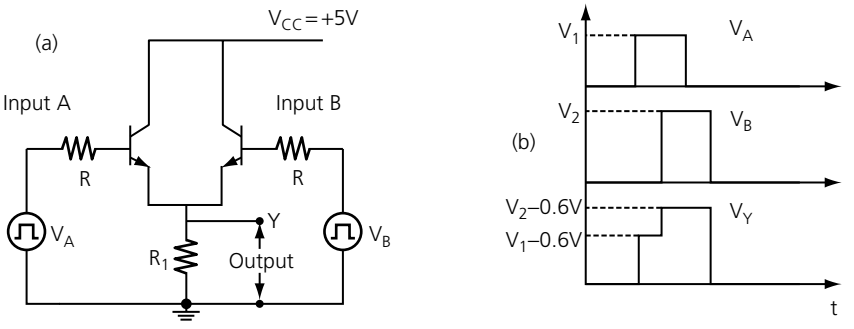


Fig. 9.30. (a) A rudimentary OR gate, (b) waveforms.

Copyright © 2005 IOP Publishing Ltd.

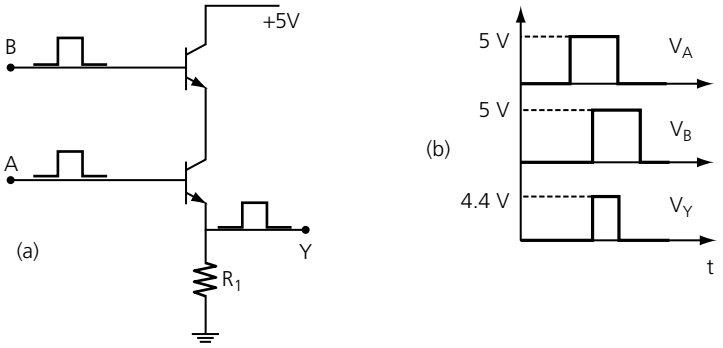


Fig. 9.31. (a) A transistor AND gate or coincidence circuit, (b) wafeforms.

above 0.6 V at either *A or B*. Waveforms are shown on figure 9.30(b), where it is assumed arbitrarily that input *B* is larger than at *A*.

AND gate

A second digital circuit is shown in figure 9.31. An input signal > 0.6 V must be present at both *A and B* for current to flow through both transistors and create an output at *Y*. If the inputs are 5 V, the output is 4.4 V. This circuit is called an AND gate or coincidence circuit. If only one input is present (or none), no current flows and the output is zero. In the commercial version, refinements are necessary to bring the output back up to 5 V.

9.7 Voltage Amplification

It is desirable to achieve voltage amplification as well as current amplification. The circuit of figure 9.32 achieves both. The output is now taken from the collector

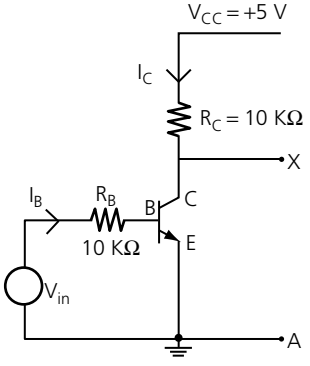


Fig. 9.32. Alternative resistor configuration giving voltage amplification.

Copyright © 2005 IOP Publishing Ltd.

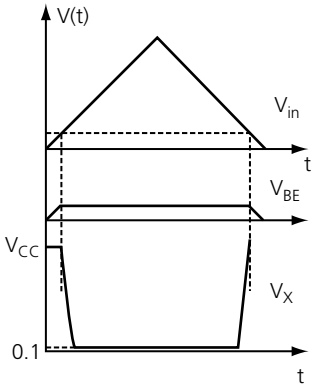


Fig. 9.33. Waveforms from the circuit of Fig. 9.32 for a triangular input.

instead of the emitter. Waveforms are shown in figure 9.33. The resistor R_B is there just to protect the diode junction BE against excessive voltage; the base voltage V_B swings up to a maximum of 0.6 V. If V_{in} is positive, $I_B = (V_{in} - 0.6)/R_B$ and $I_C = \beta I_B$. Then $V_X = V_{CC} - I_C R_C$. As I_C rises V_X falls, and it may fall as low as 0.1 V if I_B is large enough. The transistor is then **saturated**.

This time, both current and voltage are amplified. Because of the voltage amplification, the switching action is more definite than that of the emitter follower and for this reason the circuit is often preferred in digital logic. With suitable choice of R_B and R_C, input signals switch the output decisively between the supply voltage $V_{CC} = 5$ V and 0.1 V. The output is suitable for application direct to subsequent circuits; the fact that the lower level is 0.1 V instead of strictly zero does not matter because it requires 0.6 V to switch a later circuit. The difference between 0.6 V and 0.1 V is the **noise margin** for this type of circuit. The higher noise margin is one reason for preferring silicon based devices over germanium.

The *pnp* transistor has the reverse doping to *npn*. The base is *n*-type and emitter and collector are *p*-type, figure 9.34. Its operation is precisely the same as that of the *npn* transistor except that all voltages and currents are reversed in sign and except that the mobility of holes carrying the current is somewhat less than that of

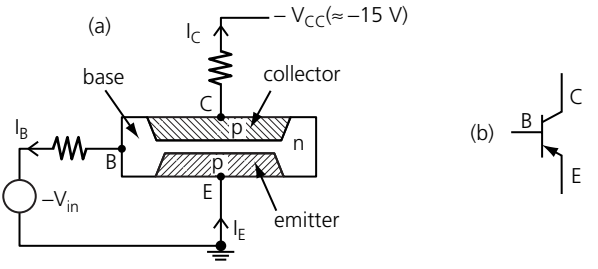


Fig. 9.34. (a) the pnp transistor and (b) its current symbol.

Copyright © 2005 IOP Publishing Ltd.

electrons in the *npn* transistor. Any circuit which is designed for *npn* transistors can equally well be implemented with *pnp* transistors by reversing all polarities.

9.8 Biasing

For digital logic the circuits of figures 9.30 and 9.31 are fine. They act as simple switches. However, suppose we want instead to amplify faithfully an AC input voltage $V = \sin \omega t$ V. In either of the circuits discussed so far the transistor comes on only when $V \geq 0.6$ V. The remedy is to superimpose the input signal on a DC level of say 2 V. Then the input swings between 3 and 1 V and the transistor stays on through the whole cycle. Figure 9.35 shows successive improvements in the way this may be achieved.

The most primitive is with the circuit of figure 9.35(a). The battery V_0 drives current through R_2 followed by the parallel combination of R_1 and the transistor; by suitable choice of resistors the DC level at the base can be held at 2 V. The AC source drives current through R_1 and the parallel combination of R_2 and the transistor. The superposition theorem gives the general picture but is not accurate because the transistor is non-linear; chapter 11 will investigate how to do the algebra properly.

In practice, separate DC voltages for V_{CC} and V_0 are inconvenient. A possible alternative is shown in figure 9.35(b). The power supply drives a current through

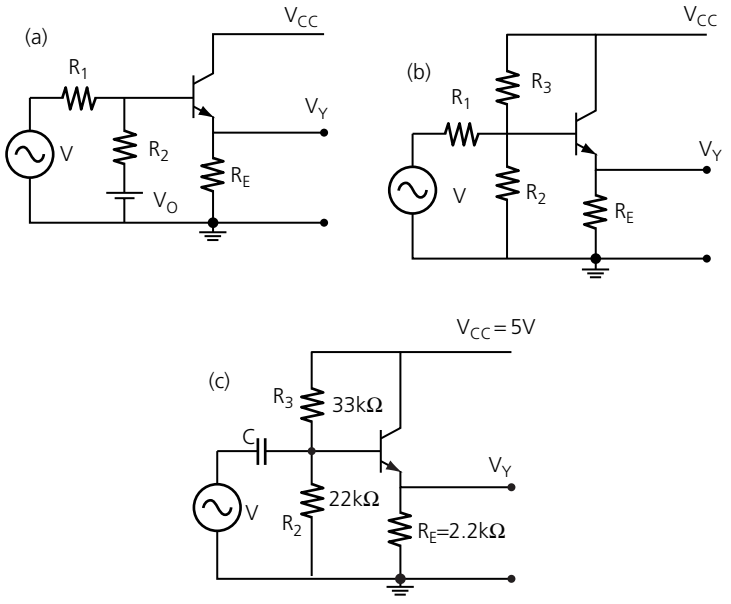


Fig. 9.35. Three ways of biasing the transistor for AC operation.

Copyright © 2005 IOP Publishing Ltd.

R_3 followed by the parallel combination of R_1, R_2 and the transistor and sets up a DC level at the base. However, the circuit has a disadvantage. The AC source V drives R_1 and the parallel combination of R_2, R_3 and the transistor. There is a loss of voltage across R_1. This resistor cannot be set to zero without shorting the base to earth.

In figure 9.35(c), R_1 is replaced by a capacitor. If this is chosen to be large, its impedance $1/j\omega C$ can be made negligible at the operating frequency so that the full AC signal is applied to the base of the transistor. It is conventional to choose R_2 and R_3 so that $\sim 90\%$ of the DC current supplied by V_{CC} flows through R_2 and 10% into the base of the transistor. Then R_1 and R_2 act as a potential divider; to achieve a 2 V level at the base requires $R_3/R_2 \simeq 3/2$. Suitable component values are shown in figure 9.35(c).

You are recommended strongly to experiment with the circuits shown in figures 9.27–35 to get an intuitive feeling for transistor operation.

9.9 Exercises

1. Write notes on the conduction of electrons in a metal and in a pn diode. Why does the current in a diode rise rapidly with applied voltage? What gives rise to current amplification in a bipolar transistor? How may this be used to amplify a voltage?

2. (QMC). In the circuit of figure 9.36, the switch is closed at time $t = 0$. By considering the Thevenin equivalent of the part of the circuit to the left of AB as a function of the voltage across AB, sketch the voltage across the inductor as a function of time. You may assume that the diode has zero forward and infinite reverse resistance, and that it switches at zero applied voltage. (Ans: For $V_{AB} > 3$ V, $V_{EQ} = 4$ V, $R_{EQ} = 25\ \Omega$; for $V_{AB} < 3$ V, $V_{EQ} = 5$ V, $R_{EQ} = 50\ \Omega$. See figure 9.37 for the waveform.)

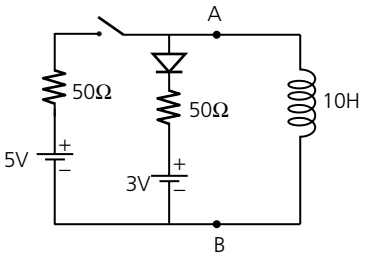


Fig. 9.36.

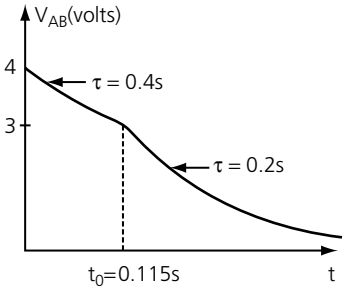


Fig. 9.37.

Copyright © 2005 IOP Publishing Ltd.

3. (QMC). In the circuit of figure 9.38, D_1 and D_2 are two identical diodes; V_A and V_B are inputs derived from square wave generators of low output impedance and having quiescent outputs at earth. At time $t = 0$, square waves $V_A = +4$ V lasting for 1 ms and $V_B = +4$ V lasting for 2 ms are applied. Describe quantitatively the form of the output V_0, assuming the diodes have a forward resistance ρ above zero applied voltage and infinite reverse resistance. (Ans: Figure 9.39.)

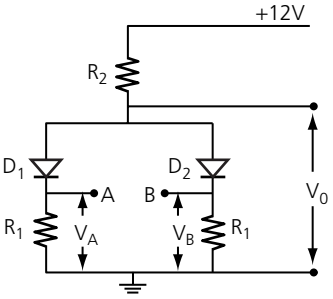
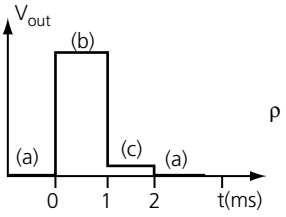


Fig. 9.38.



$$(a) = \frac{\rho/2}{\rho/2 + R_2} \times 12V$$

$$(b) = 4 + \frac{\rho/2}{\rho/2 + R_2} \times 8V$$

$$(c) = \frac{\rho \times 12}{\rho + R_2}\ V$$

Fig. 9.39.

4. (QMC). In the circuit shown in figure 9.40, the operational amplifier has a large amplification G and high input impedance. An AC voltage 5 sin ωt V is applied at point X. By considering the positive and negative parts of the cycle, find the output voltage at point Y. (Ans: Figure 9.41.)

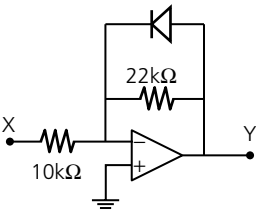


Fig. 9.40.

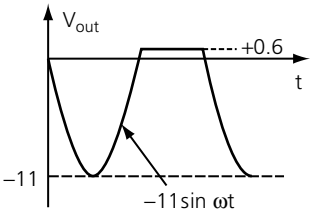


Fig. 9.41.

5. Find the peak current through a half-wave rectifier, figure 9.21(a), ignoring the forward resistance of the diode. Take the applied voltage to be 250 V RMS, $R_{out} = 200\ \Omega$.

6. What is the minimum value of the load resistor R_L in figure 9.42 if the magnitude of the ripple is not to exceed 5% of the peak voltage. What will

Copyright © 2005 IOP Publishing Ltd.

be the DC current through the load when R_L has this minimum value? (Ans: 0.295 A, 0.094 A; 800 Ω, 0.44 A.)

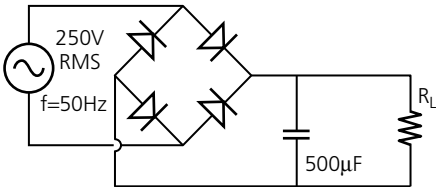


Fig. 9.42.

7. A positive voltage V_{in} is applied to the circuit of figure 9.32. Find (a) the current I_B through the base, (b) the voltage at X, assuming $I_C/I_B = 100$, and (c) the voltage amplification for AC signals, $A = \mathrm{d}V_X/\mathrm{d}V_{in}$. What happens if V_{in} is negative? (Ans: $I_B = (V_{in} - 0.6)/R_B$, $V_X = 15 - 100R_C(V_{in} - 0.6)/R_B$, $A = -100R_C/R_B = -100$; $I_B \to 0$, $V_X \to 15$ V, $A \to 0$.)

8. Suppose that in the OR gate of figure 9.30, the supply voltage V_{CC} is 5 V, $V_A = 5$ V, $V_B = 0$, $R = 10$ kΩ and $R_1 = 100$ Ω. In addition there is a resistor $R_2 = 1$ kΩ between the power supply and the collectors of the transistors. Assuming $V_{BE} = 0.6$ V and $V_{CE} = 0.1$ V when the transistor is on, find the collector current I_C, the current I_A into the base of transistor A and the voltages at emitter and collector. How do these change when (a) $R_1 = 0$, (b) $R_2 = 0$?. In (b), assume $\mathrm{I}_C = 100\,\mathrm{I}_B$. (Ans: 4.42 mA, 390 μA, 0.48 V, 0.58 V; (a) 4.9 mA, 440 μA, 0 V, 0.1 V; (b) 22 mA, 0.22 mA, 2.21 Vm 15 V.)

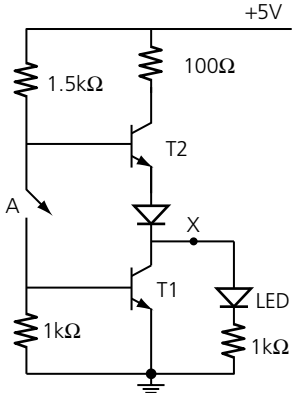


Fig. 9.43.

9*. (RHBNC). In the circuit of figure 9.43, estimate (i) the voltage at X, (ii) the current through the point X when the switch A is open and when it

Copyright © 2005 IOP Publishing Ltd.

is closed. State when the light emitting diode LED is on. (Ans: With A open, $T1$ is cut off and $T2$ on, LED on, $I_X \simeq (5 - 1.8)/1.1$ mA = 2.9 mA, $V_X \simeq 3.8$ V; with A closed, $T1$ is on, $T2$ off because $V_A \simeq 0.6$ V, $V_X \simeq 0$ V, LED off.)

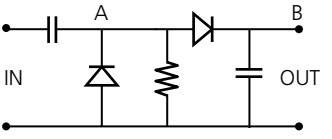


Fig. 9.44.

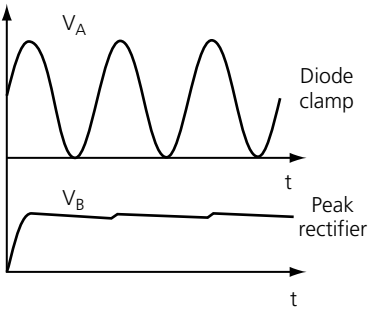


Fig. 9.45.

10*. An AC voltage $V_{in} = V_0 \sin \omega t$ is applied to the network shown in figure 9.44. Describe qualitatively the waveforms which appear at points A and B a long time after the input is switched on. (Ans: Figure 9.45.)

Copyright © 2005 IOP Publishing Ltd.

10

The Field-Effect Transistor (FET)

10.1 Gate Action

We now come to the other important class of transistor; its mode of operation is quite different to the bipolar transistor. First a rough physical picture of its operation will be constructed and then actual characteristic curves will be discussed. First let us consider the **junction FET or JFET**.

Figure 10.1(a) shows a slice of lightly doped n-type material. Suppose a voltage is applied between connections at the two ends. The material conducts and follows Ohm's law. The terminal at which electrons enter the material is called the **source** and the other terminal the **drain**.

Figure 10.1(b) shows the effect of inserting a heavily doped p-type region (called the **gate**) between source and drain. Suppose initially that it is electrically connected to the source. Carriers diffuse across the pn junction and a depletion layer develops, where free charges from the two materials have neutralised. The static charges are positive in the n-region and negative in the p-region. The depletion layer at the gate extends further into the lightly doped n-channel than into the heavily doped p-region. The channel width in the n-type material is squeezed and the current from source to drain is reduced.

Figure 10.1(c) shows the effect of applying a negative voltage V_{GS} between gate and source. Mobile electrons are repelled into the n region and the depletion layer expands. This voltage controls the width of the channel and the current I_{DS} between drain and source. It is the basis for controlling a large current by means of a small gate voltage, i.e. an amplifier.

Pinch-off

Characteristic curves are shown in figure 10.2. The essential feature of the curves at the right is that the current I_D to the drain saturates for large V_{DS}; it does so at a value controlled in a sensitive way by V_{GS}. The curve at the left shows the saturation current against V_{GS}.

Copyright © 2005 IOP Publishing Ltd.

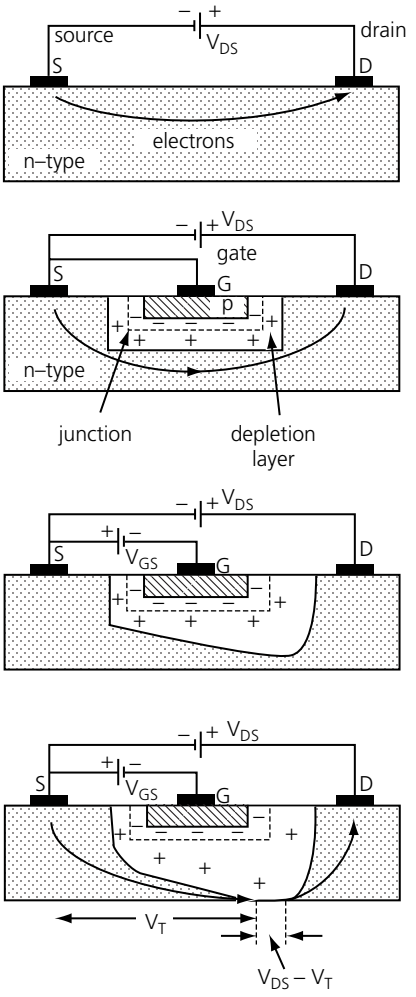


Fig. 10.1. The effect of the gate on electron flow in the n-channel, (a) no gate, (b) $V_{GS} = 0$, (c) V_{GS} small and negative, (d) large negative V_{GS}.

The depletion layer in figure 10.1(c) may be modelled using electrostatics. What emerges is that the extent of the depletion later is proportional to $V^{1/2}$, where V is the total barrier height across the junction. This barrier is the sum of the applied bias and the voltage increase along the conducting channel due to Ohm' law. For large enough negative gate voltage $(-V_T)$, the current is cut off. This is called the threshold voltage.

Next suppose the gate voltage is above this, allowing current to flow. Consider first $V_{GS} = 0$. For small V_{DS}, Ohm's law is obeyed and I_D rises linearly with V_{DS}. Along the length of the conduction channel, there is a voltage increase

Copyright © 2005 IOP Publishing Ltd.

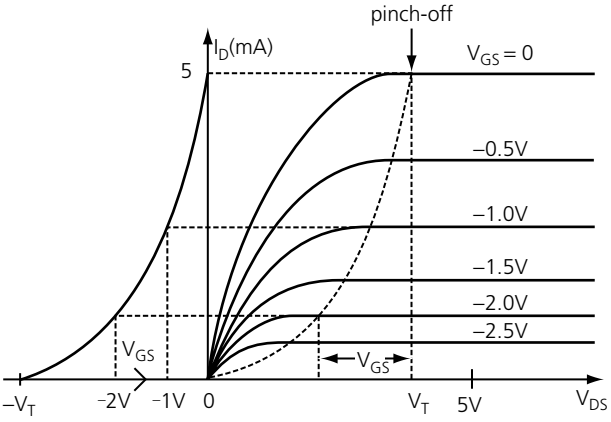


Fig. 10.2. Characteristic curves for the junction FET. The curve to the left shows the maximum (plateau) value of I_D v. V_{GS}.

from source to drain, because of the ohmic resistance of the channel. So the voltage difference between gate and channel increases from a small value near the source to a larger value near the drain. Consequently, the depletion layer is wider at the end nearer the drain, figure 10.1(c). As the depletion layer expands with increasing V_{DS}, I_D begins to flatten off. Eventually, when V_{DS} reaches the critical value V_T, the depletion layer fills the channel except for a narrow region carrying the current. This is called **pinch-off**. For values of V_{DS} above this, no further current increase occurs; the length of the narrow channel increases to compensate the increase in V_{DS}. The voltage drop $(V_{DS} - V_T)$ occurs over a short distance, figure 10.1(d).

Next consider what happens as V_{GS} is changed to say -0.5 V. This voltage reduces the channel width everywhere, so the current falls. Pinch-off is reached when the voltage in the channel is just sufficient to close it, i.e. when $V_{DS} = V_T + V_{GS}$; because V_{GS} is negative, this is a slightly lower value of V_{DS} than before. The left-hand curve on figure 10.2 shows the saturation current against V_{GS}; it is approximately parabolic:

$$I_{DS}^{sat} \simeq K(V_{GS} + V_T)^2.$$

The dashed curve on figure 10.2 showing the value of V_{DS} for pinch-off has the same parabolic shape because pinch-off occurs when $V_{DS} = V_T + V_{GS}$.

The title 'field-effect transistor' (FET) is given to all the devices considered in this chapter because the electric field of the gate controls the width of the depletion layer and hence the current flow. In the *FET*, the current in the n-region is carried by **majority carriers** (electrons). This is in contrast to the bipolar transistor, where in the controlling base region of the *npn* transistor electrons are minority carriers.

Copyright © 2005 IOP Publishing Ltd.

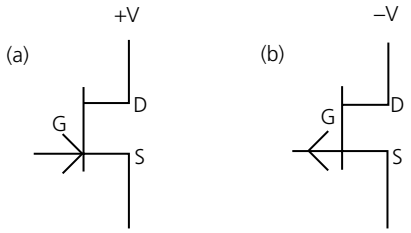


Fig. 10.3. Circuit symbols for (a) n-channel, (b) p-channel JFETs.

In figure 10.1, the arrangement of source and drain is symmetric about the gate, and the source and drain are interchangeable. However, in high speed devices, the gate is positioned close to the source, so as to reduce the capacitance between gate and drain, and hence reduce the Miller effect (Chapter 8).

The circuit symbol for the n-channel *JFET* is shown in figure 10.3(a). The arrow indicates a diode junction between p-type gate and n-type channel. A p-channel device is made using n-type gate and reversed polarities for V_{DS} and V_{GS}. Its circuit symbol is shown in figure 10.3(b). Its properties are very similar to the n-channel device, except for the reversal of polarities.

10.2 Simple FET Amplifiers

Source follower

With appropriate voltage supplies and bias resistors, the *JFET* can be used to make amplifiers in a very similar way to the bipolar transistor. Figure 10.4 shows a **source follower**, the analogue of the emitter follower. The static voltage at the gate is zero. The transistor conducts and the current through resistor R_S biases the source S to a potential above earth. Equilibrium is established when V_{GS} creates a current I_D such that $V_{GS} = -I_D R_S$. This situation is illustrated on figure 10.5. The voltage across R_S is given as a function of I_D by the straight line, known as the **load line**. The **operating point**, marked by the cross, is given by the condition that $V_{GS} + I_D R_S = 0$.

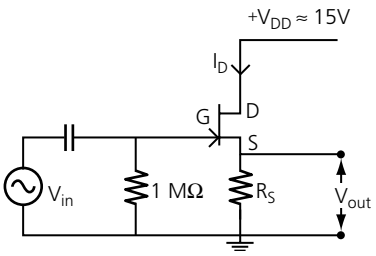


Fig. 10.4. The source follower.

Copyright © 2005 IOP Publishing Ltd.

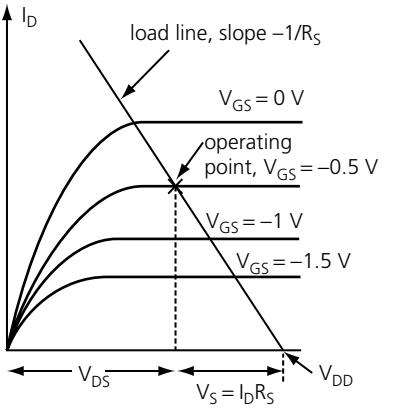


Fig. 10.5. The load line and determination of the operating point.

If an input AC signal V_{in} is now applied, the operating point moves up or down the load line. The changes ΔV_{GS} and ΔV_S may be followed graphically. However, it is easier to do so algebraically using

$$\frac{\mathrm{d}I_D}{\mathrm{d}V_{GS}} = g_m, \tag{10.1}$$

which is called the mutual conductance. Then

$$V_{in} = \Delta V_{GS} + R_S \Delta I_D = \Delta V_{GS}(1 + R_S g_m)$$

$$\Delta V_S = R_S \Delta I_D = R_S g_m \Delta V_{GS} = \frac{R_S g_m}{1 + R_S g_m} V_{in}. \tag{10.2}$$

Typical values of g_m are (2–3) $\times 10^{-3}$ mho. If $R_S \gg 1/g_m$, i.e. $\gg 500\ \Omega$, V_S follows V_{in} closely.

Common source amplifier

Figure 10.6(a) shows how to obtain voltage amplification by putting a resistor R_D between supply and drain. The voltage change at the drain is given by

$$\Delta V_D = -R_D \Delta I_D = \frac{-R_D g_m}{1 + R_S g_m} V_{in}. \tag{10.3}$$

The minus sign arises from the fact that as V_{in} goes positive I_D increases and V_D falls. It would improve the amplification if $R_S = 0$. This is not possible without upsetting the biasing of the transistor. However what can be done is to put a large capacitor across R_S, as shown by the dashed symbol in figure 10.6. The term R_S is replaced by the impedance of R_S and C in parallel and equation (10.3) becomes

$$\Delta V_D = \frac{-R_D g_m V_{in}}{1 + R_S g_m / (1 + \mathrm{j}\omega C_S R_S)}.$$

Copyright © 2005 IOP Publishing Ltd.

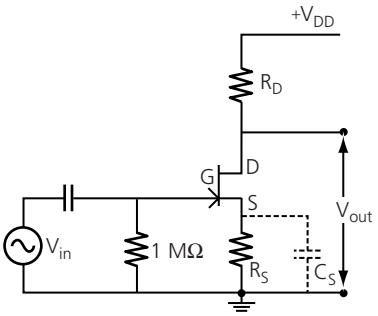


Fig. 10.6. A JFET voltage amplifier.

For AC signals with $\omega \gg 1/C_S R_S$ the denominator is now 1 for AC inputs. The large capacitor C is essentially holding the source S at a fixed DC voltage. The voltage amplification is then $-R_D g_m$.

In this simple analysis, the dependence of I_D on V_{DS} has been neglected, assuming the amplifier operates in the region where the characteristic curves are flat. The effect of $\mathrm{d}I_D/\mathrm{d}V_{DS}$ will be included in a fuller analysis in the next chapter, but actually has rather small effect. For the moment it is more important to draw attention to the fact that the characteristic curves are not equally spaced, so $g_m = \mathrm{d}I_D/\mathrm{d}V_{GS}$ is not really constant. Thus the amplification will depend on the magnitude of V_{in} and the choice of operating point. It will be larger for positive inputs than negative, because of the non-uniform spacing of the characteristic curves. This does not matter in digital circuits, where the output voltage is HIGH or LOW, and precise values of the amplification are not vital. However, to make a linear amplifier, negative feedback is required as explained in Chapter 7.

Differences from the bipolar transistor

The *JFET* differs from the bipolar transistor in two important respects. Firstly, the input resistance is very high because the gate-source junction is reverse biased; the bipolar transistor however normally operates forward biased and has a fairly low input resistance (a few kΩ). The *JFET* has an input resistance of order 1–1000 MΩ, but other versions of the *FET* we shall meet soon have an even higher input resistance, which can be as much as $10^{14}\Omega$. Here lies a problem. It is easy for charge to build up on the gate because of a large time constant CR. This charge can generate a voltage large enough to damage the gate-source junction. The purpose of the 1 MΩ resistor in figures 10.4 and 10.6 is to prevent the gate voltage from floating because of the high input resistance.

Care is needed not to expose *FETs* to situations where charge build-up may cause damage, for example by rubbing them against clothing. Even electrostatic charge on the body can cause damage, so earthed wrist-straps are advisable when handling *FETs*. The *FETs* are generally delivered in wrapping with reasonable

Copyright © 2005 IOP Publishing Ltd.

conductance, to prevent charge building up on the gate. When they are built into integrated circuits, Zener diodes are usually inserted across the inputs as a protection against large voltages of either polarity.

Secondly, the *FET* has much lower noise than the bipolar transistor. The latter has shot noise associated with the input base current. This is amplified by a factor $\beta \simeq 100$ in the collector circuit. In the *FET*, this source of noise is absent, though there is still Johnson (voltage) noise across the 1 MΩ resistor; in order to reduce noise, this resistor may be reduced. The *FET* is then usually the choice for the first stage of a low noise preamplifier.

10.3 MOSFETs

There are several variants on the physical construction of the *FET*, for example that shown in figure 10.7(a). The source and drain regions are heavily doped n-type regions; (n^+ denotes heavy doping.) They are joined by a thin channel of lightly doped n-region, but the gate is not connected directly to it. Instead, the gate is a metal layer separated by a thin ($\sim 10^3$ Å) insulating layer of SiO_2. If the gate is negatively biased, it creates an electric field which attracts holes from the p-type substrate into the n-channel region and repels electrons. This forms a depletion layer in the channel, controlling its conduction properties in a way directly similar to the junction *FET*. Such a device is called a **depletion MOSFET**, standing for Metal Oxide Semiconductor *FET*. The word 'depletion' signifies that the gate operates by depleting the n-channel. Because the gate is insulated from the n-channel, the input resistance is very high, typically 10^9 Ω

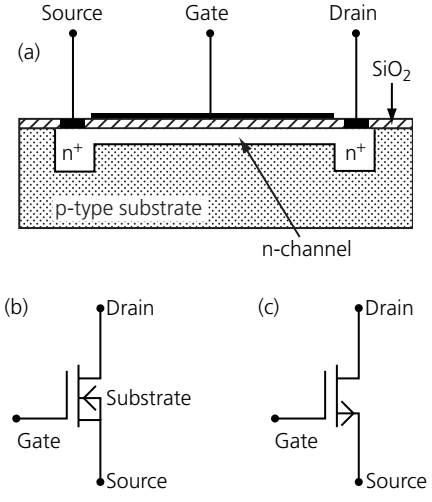


Fig. 10.7. (a) the depletion MOSFET, (b) and (c) its circuit symbols.

Copyright © 2005 IOP Publishing Ltd.

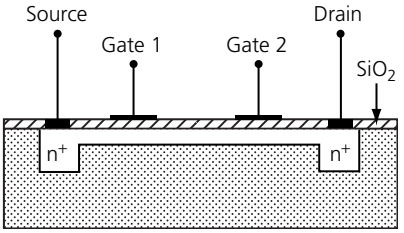


Fig. 10.8. MOSFET with two gates.

and sometimes $10^{12} - 10^{14}$ Ω. An alternative name is *IGFET*, for **insulated gate** *FET*.

Two common circuit symbols for this device are shown in figure 10.7(b) and (c). The separation of the gate symbol from the transistor clearly indicates the physical separation of the metal gate from the *n*-type channel. The arrow in figure 10.7(c) indicates the direction of current flow from drain to source. However, the official symbol shown in (b) does not indicate the current direction. Instead, the arrow indicates that the substrate is *p*-type, making a *pn* junction between substrate (*p*) and channel (*n*). But the direction in which the arrow points is a source of confusion to the beginner.

It is straightforward to fabricate depletion MOSFETs with more than one control gate at different points along the length of the *n*-channel, as in figure 10.8. This may be useful in two ways. Firstly, the gates may carry signals which independently influence the logic of the current flow. Secondly, the gate near the drain may be held at a fixed potential with respect to the source; lines of electric field from the drain mostly end on this gate. This reduces the capacitance between the first gate and the drain, and hence reduces the Miller effect. Transistors with multiple source or drain or emitter or collector are employed frequently in integrated circuits.

Figure 10.9 shows a version of the MOSFET with *no* channel between drain and source. If, however, the gate is *positively* biased, it attracts electrons to the surface region and induces an *n*-type channel just beneath the surface. Its characteristics are shown in figure 10.9(b). It is called an **enhancement** MOSFET, since the positive voltage applied to the gate enhances current flow.

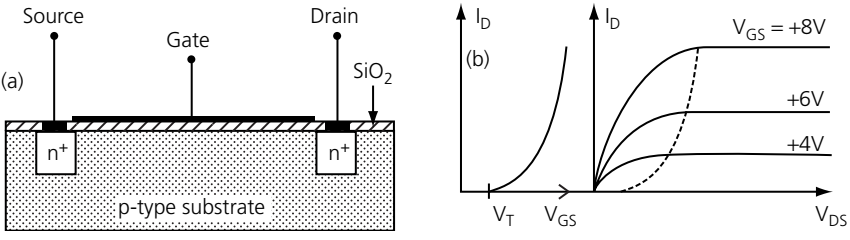


Fig. 10.9. (a) Enhancement MOSFET, (b) its characteristics.

Copyright © 2005 IOP Publishing Ltd.

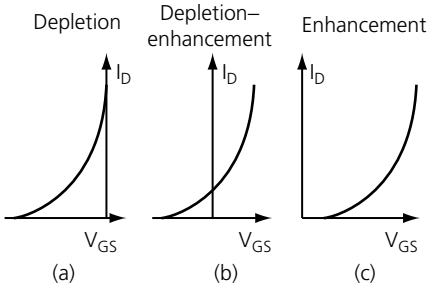


Fig. 10.10. Characteristics of (a) depletion, (b) depletion–enhancement. (c) enhancement MOSFETs.

A continuous range of possibilities between enhancement and depletion exists. Figure 10.10 shows three sets of characteristics: (a) the depletion MOSFET where the gate is negatively biased for operation, (b) the depletion-enhancement MOSFET, where the operating point is with the gate at or close to earth (in the former case making biasing unnecessary), and (c) the enhancement MOSFET.

10.4 Fabrication of Transistors and Integrated Circuits

The way in which an n-channel enhancement MOSFET may be made is indicated by the steps in figure 10.11. The process starts (a) from a lightly doped p-type substrate, typically 10 μm thick. For reasons we shall come to later, this is in turn created on an n-type substrate typically 100 μm thick.

At (b), the surface is oxidised to SiO_2 to provide an insulating barrier. At (c), the SiO_2 layer is covered with a light-sensitive material known as photo-resist. This is covered by a mask at points where the SiO_2 layer is to be etched away later. The photo-resist polymerises when illuminated, making it chemically resistant. The remaining area under the mask is dissolved at step (d); at (e) the SiO_2 is etched away chemically, exposing the source and drain regions of figure 10.9(a). The n-type regions are now formed by diffusing in donors, (f). Similar steps are then used to deposit the aluminium gate and connections at the source and drain, (g). The bipolar transistors of figure 10.12 may be made by similar processes.

Integrated circuits

It is an obvious extrapolation of the technique to make whole **integrated circuits** in the silicon rather than individual transistors. Small resistors (<1 kΩ) may simply be channels with suitable doping. Larger resistors (up to 100 kΩ) are made from MOSFET transistors with fixed gate biases which govern the resistance. Small capacitors may be created like the gate of a MOSFET, using a surface deposit of metal as one plate and a channel region as the other. Small capacitors (<50 pF)

Copyright © 2005 IOP Publishing Ltd.

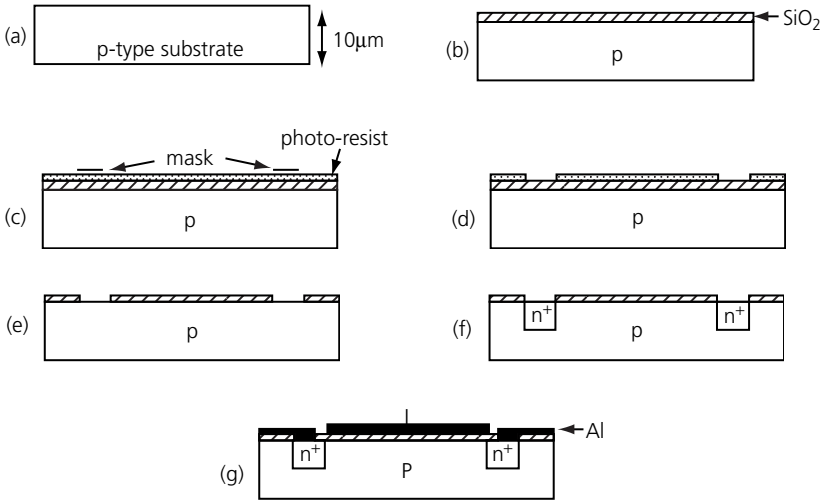


Fig. 10.11. Steps in fabricating an enhancement MOSFET.

may also be made from a *pn* junction with reverse bias. However, large capacitors are expensive in surface area and are avoided in circuit design. For this reason, most circuits are designed to be DC coupled. Inductors are impossible and are substituted by the gyrator of exercise 7.9.

In this way, whole integrated circuits are constructed. These become very cheap when mass produced, despite the capital cost of the production process. The limitations are (a) the yield of successful circuits in a batch process, (b) problems of predicting precise circuit behaviour. Progressively larger circuits with smaller components have been developed over the years. Large scale integration (LSI) uses 1000–10000 components per chip and very large scale integration (VLSI) > 10000. Computer memories are examples of VLSI circuits, where the emphasis is on massive repetition of a single simple unit. The physical dimensions of components are only a few tens of microns or less and the wavelength of light used in the masking procedure has become a real limitation. Present developments to even smaller dimensions replace optical masks by electron beam lithography.

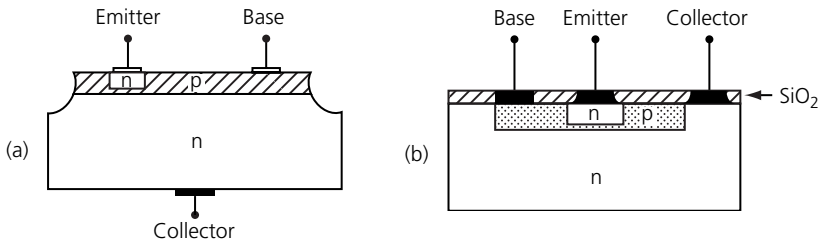


Fig. 10.12. (a) Mesa and (b) planar type bipolar transistors.

Copyright © 2005 IOP Publishing Ltd.

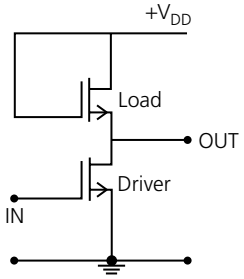


Fig. 10.13. A MOSFET inverter.

10.5 CMOS

An example of a circuit in which resistors have been eliminated is shown in figure 10.13. The load resistor is replaced by a MOSFET with its gate connected to the supply voltage. This transistor is permanently on, but by suitable choice of channel length and thickness it can be made to have any required resistance. A positive signal applied to the input switches on the lower transistor; if this is made with a lower resistance than the upper one, V_{out} falls to a low voltage.

An important development is the use of complementary MOSFETS in the same circuit. This puts to advantage the fact that n- and p-channel devices can easily be laid down on the same substrate. Figure 10.14 shows an example where the upper transistor is p-channel and the lower one n-channel. With the input zero, the upper transistor switches on and $V_{out} \simeq +V_{DD}$. With input $+V_{DD}$, the lower transistor switches on and the upper one off, so $V_{out} \simeq 0$. This circuit has the virtue that one or other transistor is off in both cases, so the DC current through them is zero, except during switching. This minimises power consumption, typically 10 nW per component.

Large-scale integrated circuits require low power dissipation and operate by using transistors to switch currents. An example of how this is achieved is shown in figure 10.15. Currents in the two arms of the circuit are shown. The transistor at the bottom of the figure acts as a constant current source, because of its fixed

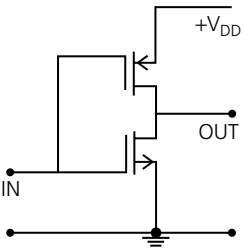


Fig. 10.14. A CMOS inverter.

Copyright © 2005 IOP Publishing Ltd.

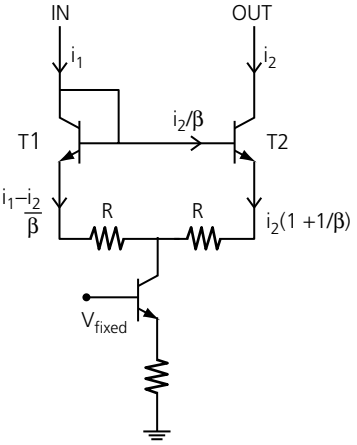


Fig. 10.15. Current Switching.

base voltage. Consequently $I_1 - I_2/\beta + I_2(1 + 1/\beta)$ is constant. This requires $I_1 = -I_2$. Any change in the current in one arm is reflected in the other. This is one version of the **Current Mirror.**

Copyright © 2005 IOP Publishing Ltd.

11

Equivalent Circuits for Diodes and Transistors

11.1 Introduction: the Diode

For small AC signals, diodes and transistors may be represented by Thevenin or Norton equivalent circuits. This makes it easy to analyse their performance. The ideas will be introduced using the diode as an example.

Suppose a DC voltage V_0 is applied to a resistor and diode in series, figure 11.1(a). From Kirchhoff's voltage law, the voltage V_d across the diode is given by

$$V_0 = I_0 R + V_d(I_0). \tag{11.1}$$

The DC resistance of the diode, V_d/I_0, varies with V_d. It does not obey Ohm's law. Using the diode as a switch, this is very obvious: its resistance is small for forward bias and very large for reverse bias.

The values of V_d and I_0 may be determined as in figure 11.1(b) from the intersection of the straight line $V_0 - IR$ with the diode curve. This intersection at $I = I_0$ is the operating point. The straight line gives the DC voltage across the diode after allowing for the potential drop across the load R.

Next consider figure 11.1(c), where a small AC voltage $v \sin \omega t$ is added to the circuit. On figure 11.2, the voltage moves between the two straight lines. The current moves up and down the characteristic curve. It changes by a small amount $\pm i$ and the voltage across the diode changes by $i\rho$, where ρ is the slope of the diode curve $\mathrm{d}V_d/\mathrm{d}I$ at $I = I_0$. Then

$$V_0 + v \sin \omega t = (I_0 + i)R + V_d(I_0) + i\rho. \tag{11.2}$$

Subtracting (11.1),

$$v \sin \omega t = iR + i\rho. \tag{11.3}$$

Copyright © 2005 IOP Publishing Ltd.

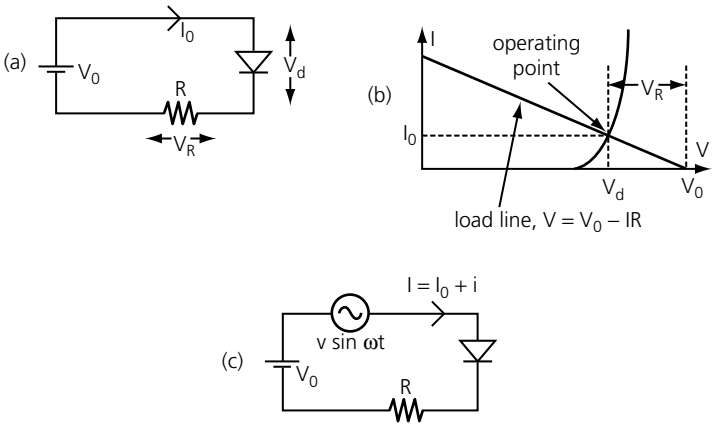


Fig. 11.1. (a) Diode circuit, (b) determination of the operating point, (c) the effect of a small AC voltage.

Equation (11.3) may be represented by means of the **AC equivalent circuit** shown in figure 11.3. It describes the AC part of the behaviour. The quantity ρ is called the slope resistance:

$$\boxed{\rho = (\mathrm{d}V_d/\mathrm{d}I)_{I=I_0}.}$$

Effectively, the behaviour of the circuit has been separated into a *DC condition* given by equation (11.1) and an *AC condition*, equation (11.3). This works even for power dissipation:

$$\overline{P} = \overline{VI} = \overline{(V_0 + v\sin\omega t)(I_0 + i\sin\omega t)} = V_0 I_0 + \tfrac{1}{2}vi.$$

The cross-terms involving $\sin\omega t$ average to zero.

The conclusion from figure 11.3 is that, as far as AC signals are concerned, we can

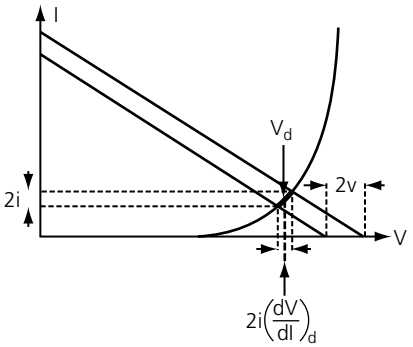


Fig. 11.2. Effect of $v\sin\omega t$.

Copyright © 2005 IOP Publishing Ltd.

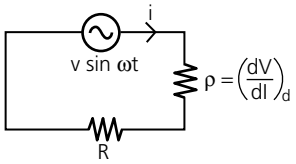


Fig. 11.3. Diode AC equivalent circuit for small signals.

(a) omit the battery (short circuit it),
(b) replace the diode by an equivalent resistance ρ.

If the diode characteristic is very steep, is ρ large or small? Can you understand the equivalent circuit direct from figure 11.2 in this case?

In Chapter 9, an algebraic expression was derived for the current in the diode:

$$I = I_0 \left[\exp\left(\frac{eV}{kT}\right) - 1 \right]. \tag{9.7}$$

Then

$$\frac{\mathrm{d}I}{\mathrm{d}V} = I_0 \frac{e}{kT} \exp\left(\frac{eV}{kT}\right) \simeq \frac{eI}{kT}$$

when V is well away from zero and the 1 in (9.7) can be neglected. Hence the slope resistance of the diode is given by

$$\boxed{\rho = \frac{\mathrm{d}V}{\mathrm{d}I} \simeq \frac{kT}{e}\frac{1}{I} \simeq \frac{1}{40}\frac{1}{I(\mathrm{A})} = \frac{25\ \Omega}{I(\mathrm{mA})}.} \tag{11.4}$$

This is a useful rule of thumb. It applies also to the diode action between base and emitter in a bipolar transistor.

If the AC signal v is not infinitesimally small, ρ varies as the diode goes up and down the characteristic. Even in this case an average value of ρ in figure 11.3 may give a useful approximation.

From equation (11.4), what change in V do you expect as I varies from 1 to 10 mA*? Use* $\mathrm{d}V = \rho \mathrm{d}I$.

11.2 An Equivalent Circuit for the Bipolar Transistor

As a reminder, the typical characteristic curves for a bipolar transistor are reproduced in figure 11.4. It is possible, but laborious, to follow the behaviour of a transistor graphically, direct from these curves. However, when dealing with small AC signals it is easier to replace the transistor by an equivalent circuit, as for the diode.

Copyright © 2005 IOP Publishing Ltd.

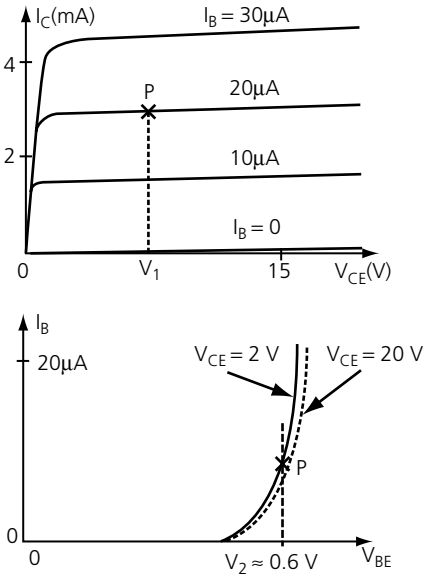


Fig. 11.4. Characteristics of the bipolar transistor.

The derivation of the equivalent circuit is slightly more complicated than for the diode because collector current I_C now depends on two variables, I_B and V_{CE}, figure 11.4(a). So does V_{BE}, figure 11.4(b). A static operating point P is defined by DC voltages applied between emitter and collector ($V_{CE} = V_1$ on figure 11.4) and between emitter and base ($V_{BE} = V_2$). Then the dependence on small variations of both I_B and V_{CE} gives the equivalent circuit.

Output side

Consider first the output (collector) circuit. We can get from point P of figure 11.5(a) to any neighbouring point Q by first moving along one of the curves with I_B fixed, then changing I_B keeping V_{CE} fixed. In the first step, as V_{CE} varies, figure 11.5(b) shows that

$$\delta I_C = \left(\frac{\partial I_C}{\partial V_{CE}}\right) \delta V_{CE}.$$

The ∂ notation indicates that only these two variables I_C and V_{CE} are changing; I_B is held constant. It is called a partial differential. The slope of the line on figure 11.6(b) at the point P gives the required differential.

Secondly, varying I_B, the change δI_B on figure 11.6(c) gives

$$\delta I_C = \left(\frac{\partial I_C}{\partial I_B}\right) \delta I_B.$$

Copyright © 2005 IOP Publishing Ltd.

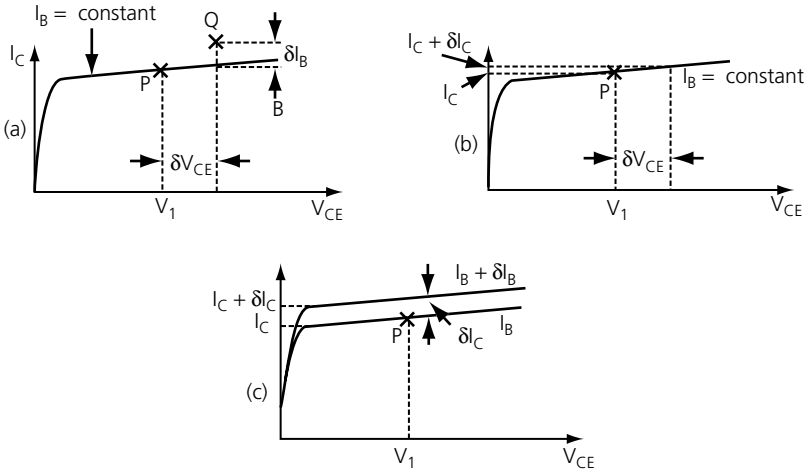


Fig. 11.5. Changes to I_C.

This time, V_{CE} is held constant. These are the only two changes to be made; varying V_{BE} reflects itself in changing I_B and V_{CE}, since it depends only on these two variables.

Putting these two terms together,

$$\delta I_C = \left(\frac{\partial I_C}{\partial I_B}\right)\delta I_B + \left(\frac{\partial I_C}{\partial V_{CE}}\right)\delta V_{CE}. \tag{11.5}$$

The notation may be abbreviated by using small letters $v_{CE} = \delta V_{CE}$, $i_C = \delta I_C$, and $i_B = \delta I_B$ for the changes,

$$\boxed{i_C = h_{fe} i_B + h_{oe} v_{CE}.} \tag{11.6}$$

Here $h_{fe} = (\partial I_C/\partial I_B)$ and $h_{oe} = (\partial I_C/\partial V_{CE})$.

This achieves the objective of deriving an equivalent circuit for the output side of the transistor. The right-hand side of figure 11.6(a) shows the Norton equivalent circuit up to terminals CE. It is described by equation (11.6) where i_C, i_B and v_{CE} are small changes in these currents and voltages. The current i_C is made up of two components, (a) $h_{fe} i_B$ from the current source, and (b) $h_{oe} v_{CE}$ through the impedance $1/h_{oe}$. The manufacturer of the transistor supplies graphs of h_{fe} and h_{oe}. However, we already know from Chapter 9 that $I_C \simeq \beta I_B$, so $h_{fe} = \partial I_C/\partial I_B \simeq \beta \simeq 100$. The f of h_{fe} stands for 'forward', denoting that a change in I_B propagates forward to a change in I_C.

The transistor characteristic of figure 11.4(a) shows rather little dependence on V_{CE} above the knee. This implies that $h_{oe} = \partial I_C/\partial V_{CE}$ is small and $1/h_{oe}$ is large. A typical value is 20 kΩ. In the majority of electronic circuits, this is considerably greater than the load placed across the output terminals. If so, $1/h_{oe}$ can be neglected and the equivalent circuit of the output becomes very simple: just a current source βi_B amplifying the base current i_B.

Copyright © 2005 IOP Publishing Ltd.

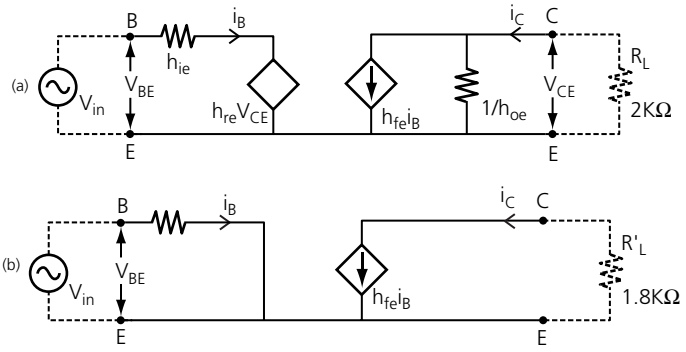


Fig. 11.6. (a) The complete equivalent circuit in the common emitter configuration, (b) a common simplification.

Input side

The input circuit can be treated likewise, expressing changes in V_{BE} in terms of δI_B and δV_{CE}:

$$\delta V_{BE} = \left(\frac{\partial V_{BE}}{\partial I_B}\right)\delta I_B + \left(\frac{\partial V_{BE}}{\partial V_{CE}}\right)\delta V_{CE} \tag{11.7}$$

or

$$\boxed{v_{BE} = h_{ie}i_B + h_{re}v_{CE}} \tag{11.8}$$

where $h_{ie} = (\partial V_{BE}/\partial I_B)$ and $h_{re} = (\partial V_{BE}/\partial V_{CE})$ are again partial differentials. In the former, V_{CE} is constant, and in the latter I_B.

The equivalent circuit for equation (11.8) is shown at the left of figure 11.6. The characteristics of figure 11.4 depend very little on V_{CE}, so h_{re} is usually negligible. In this case, the equivalent circuit of the input becomes very simple: just an input resistance h_{ie}. It is common to refer loosely to h_{ie} as the input impedance. Strictly speaking this is incorrect, since the input impedance is

$$r_{in} = v_{BE}/i_B = (h_{ie}i_B + h_{re}v_{CE})/i_B = h_{ie} + h_{re}v_{CE}/i_B.$$

In the parameter h_{re}, the r denotes 'reverse', because h_{re} is input voltage divided by output: $\partial V_{BE}/\partial V_{CE}$. The effect of h_{re} is to introduce some small feedback from output v_{CE} to input.

Suppose the circuit feeds a load R_L. Typically this is 2kΩ. If h_{re} is neglected, the complete equivalent circuit becomes figure 11.6(b), where R'_L is the parallel combination of R_L and $1/g_{oe}$. Let us now follow through the amplification of a small AC input. Suppose $v_{BE} = 10\sin\omega t$ mV, $h_{ie} = 1$ kΩ, $h_{fe} = 100$ and $1/h_{oe} = 20$ kΩ Then the base current changes by

$$\delta I_B = i_B = v_{in}/h_{ie} = 10\sin\omega t\ \mu\text{A}. \tag{11.9}$$

Copyright © 2005 IOP Publishing Ltd.

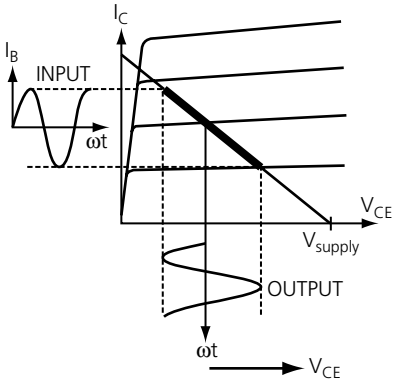


Fig. 11.7. Inversion of the input signal in the common emitter configuration.

The collector current changes by

$$\delta I_C = i_C = h_{fe} i_B = h_{fe} v_{in}/h_{ie} = \sin \omega t \text{ mA}. \tag{11.10}$$

The transistor is behaving as a current amplifier with gain $h_{fe} = 100$ and input impedance $h_{ie} = 1\ \text{k}\Omega$. The output voltage is

$$\delta V_{CE} = v_{CE} = -R'_L i_C = -1.8 \sin \omega t \text{ V}$$

and the voltage gain of the circuit (called the common emitter amplifier) is

$$G_V = -R'_L h_{fe}/h_{ie} = -180.$$

The minus sign indicates that the transistor inverts the input signal: if v_{BE} is positive, v_{CE} is negative, because of the direction of current flow through the load R'_L (from E to C). Superposing waveforms on the characteristic curves, as in figure 11.7, illustrates this inversion explicitly.

Finally, let us return to the effect of $h_{re} v_{CE}$. Suppose its magnitude is $\sim 5 \times 10^{-4} v_{CE} = -0.9 \sin \omega t$ mV. Its sign is such that in figure 11.6(a) the top of the generator $h_{re} v_{CE}$ is negative with respect to E when v_{BE} is positive. This increases the current i_B through h_{ie} by 9%. In turn this increases i_c and v_{CE} by 9%. However, this correction is likely to be within the tolerances with which h_{ie} is known. Below, we shall find that h_{ie} varies quite rapidly with operating point. So effects of h_{re} can usually be ignored.

Common base and common collector configurations of the transistor are also used frequently. A similar analysis leads to h equivalent circuits identical in appearance to the one for the common emitter, except that h parameters are labelled with a different subscript, e.g. h_{ib} for the input resistance in the common base configuration.

The manufacturer of a transistor normally supplies h parameters for these other configurations. However, this is not strictly necessary, since the performance of the transistor can be modelled in a way which allows the new h parameters to be derived from the ones we already have; let us now do this.

Copyright © 2005 IOP Publishing Ltd.

11.3 The Hybrid-π Equivalent Circuit

Figure 11.8(b) shows an equivalent circuit derived from the physical characteristics of the bipolar transistor. The resistance r_π is obtained from the slope resistance of the *pn* junction between base and emitter:

$$r_\pi = \frac{\mathrm{d}V_{BE}}{\mathrm{d}I_B} = \frac{\mathrm{d}I_E}{\mathrm{d}I_B}\frac{\mathrm{d}V_{BE}}{\mathrm{d}I_E} = \frac{(\beta+1)25\ \Omega}{I_E(\mathrm{mA})}. \tag{11.11}$$

It is also necessary to add a further resistance r_b allowing for the resistance of the terminal connection and the ohmic resistance of the base region, typically (50 Ω).

How much does r_π change as I_E goes from 1 to 10 mA *if $\beta = 100$?*

Figure 11.8(b) gives

$$v_{BE} = (r_b + r_\pi)i_B.$$

Comparing this with equation (11.8),

$$h_{ie} = r_b + r_\pi \tag{11.12}$$

$$h_{re} = 0. \tag{11.13}$$

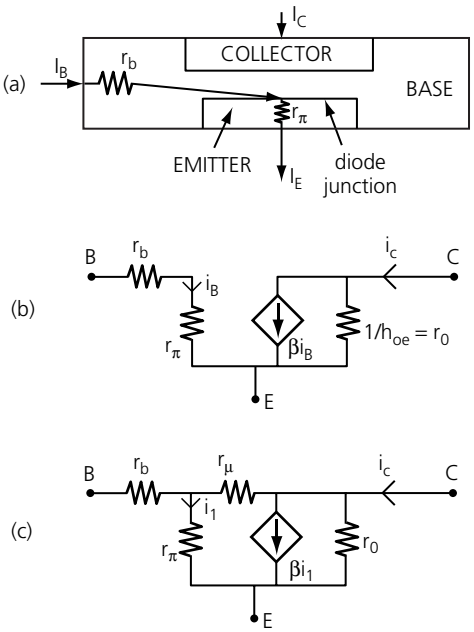


Fig. 11.8. (a) The bipolar transistor, (b) the hybrid–π equivalent circuit in simplified form, (c) including r_μ.

Copyright © 2005 IOP Publishing Ltd.

If $I_E = 2$ mA and $\beta \simeq 100$, h_{ie} and r_π are of order 1 kΩ. This is a typical figure for the input resistance of the common emitter amplifier. However, if I_E varies over the cycle of an AC signal by a significant amount r_π changes, consequently so do h_{ie} and the voltage gain. The output signal is no longer a true replica of the input. Thus the rudimentary common emitter amplifier suffers from distortion.

From equation (11.13), it is clear that figure 11.8(b) does not allow for the small dependence of the input characteristic on V_{CE}. To accommodate this, a further resistor r_μ may be added, as in figure 11.8(c) between C and B. In practice this resistor is very large (typically 1 MΩ) and makes the algebra messy. Both h_{re} and r_μ will be ignored here for simplicity. Then the output circuit between collector and emitter is identical to the h equivalent circuit of the common emitter configuration with $r_0 = 1/h_{oe} \simeq 20$ kΩ.

This model works for *all* configurations: common emitter, common base and common collector. Using this model, relations may be found between the h parameters in the different configurations.

11.4 The FET

The derivation of an equivalent circuit for the FET (of any variety) follows the same procedure as for the bipolar transistor. As a reminder, typical characteristic curves are repeated in figure 11.9. DC voltages applied at the gate and drain define an operating point P, and the equivalent circuit describes small variations around this point:

$$\delta I_D = \left(\frac{\partial I_D}{\partial V_{GS}}\right)\delta V_{GS} + \left(\frac{\partial I_D}{\partial V_{DS}}\right)\delta V_{DS} \tag{11.14}$$

or

$$\boxed{i_D = g_m v_G + (1/r_d)v_D} \tag{11.15}$$

where i and v stand for small changes in current or voltage. The quantity $g_m = \partial I_D/\partial V_{GS}$ is called the **mutual conductance**; a typical value is 2–3 mA/volt. The quantity $r_d = \partial I_D/\partial V_{DS}$ is called the drain resistance and is typically 20–100 kΩ.

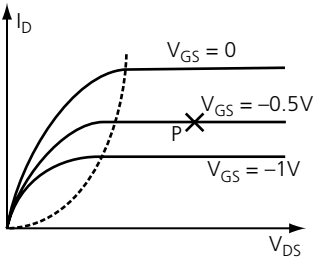


Fig. 11.9. Characteristic curves of the FET.

Copyright © 2005 IOP Publishing Ltd.

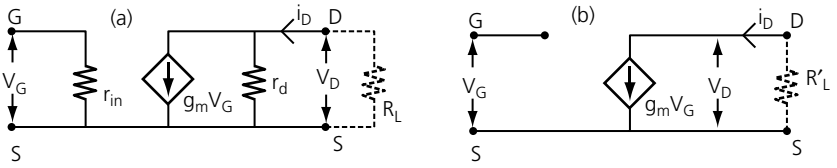


Fig. 11.10. (a) FET equivalent circuit, (b) a common simplification.

The right-hand half of figure 11.10(a) shows the equivalent circuit corresponding to equation (11.15). On the left-hand side, the input between gate and source is a reverse-biased diode with a very high resistance $r_{in} (\geq 10^6\ \Omega$, often *very* much greater). Usually this resistance is so high compared with other resistances in the circuit that the input of the FET can be treated as an open circuit, figure 11.10(b).

The equivalent circuit is very simple. It is that of a current source controlled by the input voltage. It has high input resistance and voltage gain $-g_m R'_L$, where R'_L is the parallel combination of the load R_L with r_d. Very frequently, r_d is much larger than R_L and can be neglected. The input current is very small, so the current gain is extremely high, and so is the power gain.

11.5 The Common Emitter Amplifier

Circuits used in today's integrated circuits are too complex to study here and involve tricks which depend on delicate control of doping levels. It is still useful to understand the principles of simple amplifiers which can be tried in the laboratory using a single transistor.

Figure 11.11 shows the full layout of a Common Emitter Amplifier with typical resistors used for setting the operating point. First we shall work through the DC conditions.

Suppose the supply voltage $V_{CC} = 15$ V. Next suppose the output voltage is to have the latitude to swing over a range of ± 5.5 V. (The reason for not aiming

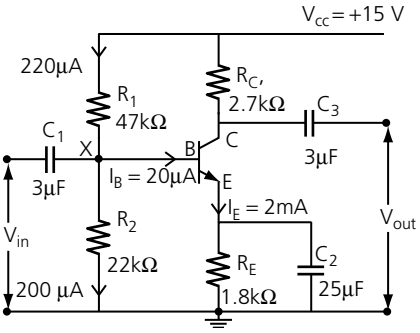


Fig. 11.11. Common emitter amplifier.

Copyright © 2005 IOP Publishing Ltd.

for ±7.5 V will appear shortly.) This choice implies setting the DC voltage of the collector at 9.5 V. It can then swing up to the supply voltage and down to 4 V. If the latter takes the collector down to the knee of the characteristic curves ($V_{CE} \simeq 0.4$ V), the DC value of the emitter voltage is $V_E \simeq 3.6$ V and the base voltage is $V_B \simeq 4.2$ V. The DC voltage V_{CE} across the transistor is then 5.9 V. There is obviously some freedom of choice in these conditions.

The choice of resistor values goes as follows. The potentiometer R_1 and R_2 fixes the DC value of the base voltage V_B and hence V_E. The value of R_E then defines the DC current I_E through the emitter and hence $I_C = \beta I_E/(\beta + 1)$ and $I_B = I_E/(\beta + 1)$. It is this requirement (that R_E sets I_E) which determines that the DC value of V_E be non-zero and prevents the output voltage from swinging all the way down to zero. The load resistor R_C defines the collector voltage, hence the DC voltage V_{CE} across the transistor.

Values of R_1 and R_2 are chosen so that 90% of the current through R_1 flows through R_2 and only 10% into the base of the transistor. As we shall see, this protects the operating point against variations in β of the transistor. The large capacitor C_2 holds the DC voltage of the emitter nearly constant, but bypasses R_E for AC signals; for the latter, the emitter is effectively earthed and the input signal is developed between emitter and base; the output signal is then developed between emitter and collector. Hence this is the common emitter configuration for AC signals.

Suppose I_E is chosen to be 2 mA. Then $V_E = 3.6$ V requires $R_E = 1.8$ kΩ. Then $R_C = 2.7$ kΩ gives $V_C = 9.6$ V. Suppose next that the transistor has a current gain of $\beta = 100$, so that $I_B = 20\mu$A and $I_1 = 220\mu$A. If the difference in current through R_1 and R_2 is ignored and if V_B is to be 4.2 V, $R_2 \simeq 19$ kΩ and $R_1 \simeq 49$ kΩ. To allow for the fact that I_B flows through R_1 but not R_2, R_2 is rounded up to 22 kΩ and R_1 down to 47 kΩ, the nearest available common values.

What change is there in operating point if all resistors are halved?

Suppose R_2 were omitted completely and V_B were set via the potential drop through R_1 due to base current I_B:

$$V_B = V_{CC} - R_1 I_B.$$

Now V_B becomes directly sensitive to I_B hence β, so this is a poor choice.

11.6 Performance of the Common Emitter Amplifier

What is of interest is the amplification of AC signals. To follow this, the transistor is replaced by its AC equivalent circuit. If the full complexity of the circuit is retained, the result is shown in figure 11.12. The way R_1 and R_C have been drawn requires comment. It has been assumed that the battery supplying power has zero impedance. Then from the point X of figure 11.11 there is a path up through R_1

Copyright © 2005 IOP Publishing Ltd.

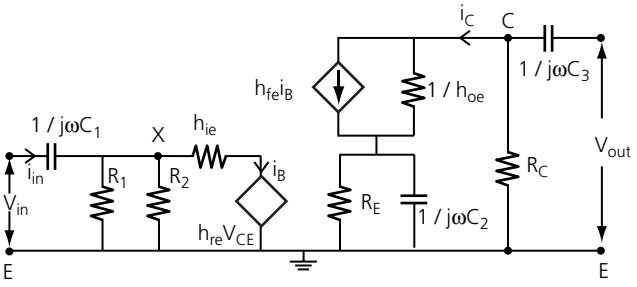


Fig. 11.12. Full equivalent circuit of the common emitter amplifier.

and through the battery to earth. (Remember that for AC equivalent circuits DC voltage sources are omitted.) Likewise, from the collector there is a path through R_C then the battery to earth.

What changes are there if the power supply has resistance R_0?

Analysing Figure 11.12 in full is possible but very laborious. It is better to make approximations that (a) $h_{re} = 0$ and (b) the capacitors have negligible impedance; at the end, these approximations can be checked and improved if necessary by back-substitution. Using these approximations the circuit simplifies dramatically to that of figure 11.13.

Input and output impedances

The interpretation of this circuit is simple. The input impedance r_{in} is that of h_{ie}, R_1 and R_2 in parallel. If h_{ie} = 1 kΩ and R_1 and R_2 have the values chosen above (22 kΩ and 47 kΩ), the latter two carry little current and the input impedance is only 6.3% less than h_{ie}. Likewise i_B is 6.3% less than i_{in}.

The output circuit is equally simple: a current generator $h_{fe}i_B$ with an output impedance r_{out} given by R_C in parallel with $1/h_{oe}$. If the latter is 20 kΩ and $R_C = 2.7$ kΩ, the output impedance R_{out} is 12% less than R_C, i.e. 2.38 kΩ.

In summary, the bias resistors play only a small role in the performance of the common emitter amplifier providing their values are chosen so that R_1 and R_2 are large compared with h_{ie} and R_C is small compared with $1/h_{oe}$.

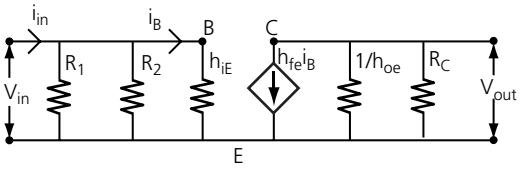


Fig. 11.13. Simplified (approximate) equivalent circuit.

Copyright © 2005 IOP Publishing Ltd.

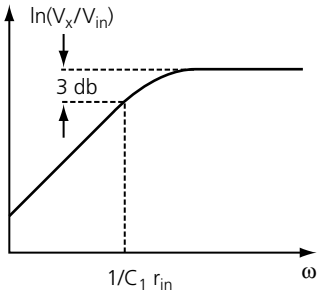


Fig. 11.14. Low-frequency cut-off due to C_1.

Current and voltage gain

If $h_{fe} = 100$, the output current $h_{fe}i_B$ is $100i_B \simeq 94i_{in}$. How much of this is delivered to the external load depends on the load impedance. For maximum current gain, the load should be small compared to R_C and the current gain of the amplifier is then 94. On the other hand, for maximum voltage gain, it should have a large impedance. The voltage gain is then $h_{fe}i_B r_{out}/h_{ie}i_B = 100 \times 2380/1000 = 238$.

If R_E and R_C are doubled, what is the effect on h_{iE} and the resulting change in the voltage gain?

Choice of capacitor values

Next, what is the effect of C_1 and C_3? The input voltage to the transistor (figure 11.12) is

$$v_X = \frac{v_{in}r_{in}}{r_{in} + 1/\mathrm{j}\omega C_1}.$$

This falls off at low frequency as shown in figure 11.14. The gain falls by 3 db (i.e. a factor $1/\sqrt{2}$) when $\omega C_1 r_{in} = 1$ in the denominator, i.e. $\omega = 1/C_1 r_{in}$. Remember that $r_{in} \simeq h_{ie}$. For a cut-off at $\omega = 2\pi \times 50$ rad/s say, a value $C_1 \simeq 3\mu$F is required. Larger values give a lower cut-off frequency. Likewise C_3 and the output impedance r_{out} of the amplifier give a low frequency cut-off at $\omega = 1/C_3 r_{out}$ and since r_{out} is not very different from r_{in}, C_3 should be roughly similar to C_1.

AC load line

The proper discussion of C_2 awaits the discussion of the emitter follower. Suppose, however, it is large enough to bypass R_E of figure 11.12. The DC load line on figure 11.15 shows the relation between I_C and V_{CE} for changes in the DC operating point. It has a slope of $-1/(R_C + R_E)$ because these two resistors appear in series

Copyright © 2005 IOP Publishing Ltd.

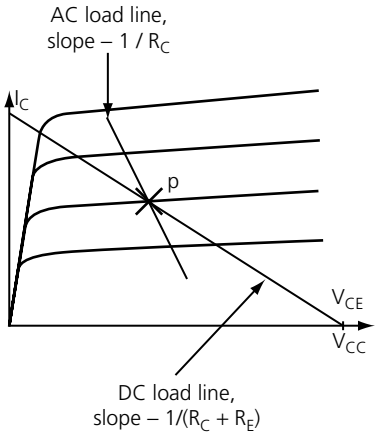


Fig. 11.15. AC and DC load lines.

with the transistor between supply voltage and earth. AC signals, however, bypass R_E. The transistor therefore responds to an AC input voltage by changes i_C and v_{CE} along the AC load line with slope $-1/R_C$.

It is recommended that you apply a small AC voltage to the circuit of figure 11.11 and follow the signal and the DC voltages with an oscilloscope. If you increase the input signal you will see distortion when the output voltage is limited either by the power supply or by bottoming. What component change eases this problem? Vary the frequency and identify by trial and error which capacitor is responsible for the low frequency cut-off. You can check the output impedance in the middle of the passband by putting load resistors across the output to halve the voltage. You can measure the input impedance likewise by putting resistors in series with the input to halve the output of the amplifier.

Feedback

The input impedance h_{ie} of the transistor is very sensitive to I_E, hence operating point, and so is the voltage gain. To remedy this, negative feedback may be used to stabilise the gain, as described in Chapter 7. One very simple way of achieving this is to move R_1 of figure 11.11 to provide a feedback path from collector to base. Because the amplifier inverts the input signal, this provides negative feedback. The value of R_1 is changed to provide the right biasing condition for the base and R_3 is chosen to provide the required amplification. The voltage gain of the amplifier $-G/(1 + GB)$ becomes approximately $-1/B = -R_1/R_3$, providing this is much less than the voltage gain G of the transistor; it can be adjusted by varying R_3.

What are then the input and output impedances?

Copyright © 2005 IOP Publishing Ltd.

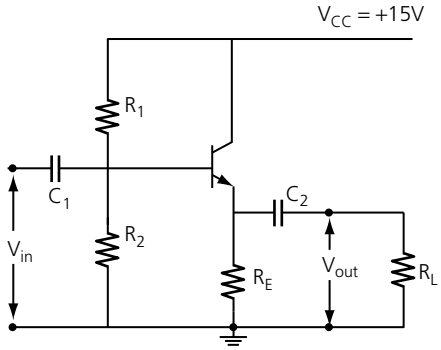


Fig. 11.16. The emitter follower.

11.7 Emitter Follower

A second amplifier used very commonly is the emitter follower. The layout of the circuit, shown in figure 11.16, is similar to the common emitter amplifier, but (a) R_C is omitted and (b) R_E does not have a bypass capacitor. The DC bias conditions are the same as for the common emitter amplifier, except that $V_{CE} = V_{CC} - V_E$.

The AC operation is very simple. The base-emitter junction acts as a diode, so $V_E \simeq V_B - 0.6$ V. Consequently, the output follows the input, except for a DC drop of 0.6 V, which has no effect on the AC behaviour. Hence for AC signals $v_E = v_B$, and the voltage gain is $\simeq 1$. There is, however, a current gain since $I_E = I_B + I_C = (\beta + 1)I_B$.

For AC signals, the battery behaves as a short circuit. Hence, as far as AC signals are concerned, the collector is at earth. Then the input signal is developed between base and collector and the output signal between emitter and collector. It is therefore the common collector amplifier.

It will emerge that the output impedance of the circuit is low, typically 20 Ω, and the input impedance r_{in} of the transistor $\simeq (\beta + 1)R_E$. Typical values of R_E are $\leq$2 kΩ. The input impedance is high and R_1 and R_2 are generally made rather larger than r_{in}, so that most of the AC current flows into the transistor; typical values of R_1 and R_2 are $\geq$ 50 kΩ The output impedance $\ll$ input impedance,

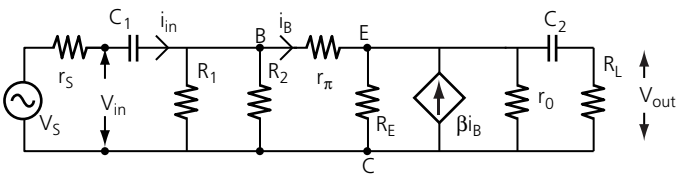


Fig. 11.17. AC equivalent circuit of the common emitter amplifier.

Copyright © 2005 IOP Publishing Ltd.

so the emitter follower is often used to drive a low impedance load from a high impedance source.

An h equivalent circuit can be substituted for the transistor, but it is more instructive to use the hybrid-π. In figure 11.17, R_1 is connected from base to earth via the battery, otherwise the circuit is obvious. In order to simplify the analysis, the approximations will be made of neglecting the impedances of C_1 and C_2. Then the load resistor R_L appears in parallel with R_E and r_0. Suppose this parallel combination has impedance R'_E.

Gains

With these simplifications, the current through R'_E is $i_E = (\beta + 1)i_B$ and the voltage gain is

$$\frac{v_{out}}{v_{in}} = \frac{i_E R'_E}{i_B r_\pi + i_E R'_E} = \frac{R'_E}{R'_E + r_\pi/(\beta + 1)}. \tag{11.16}$$

If $r_\pi/(\beta + 1) \ll R'_E$, the voltage gain is $\sim$1, as was found in qualitative fashion in the previous section.

Input impedance

In deriving the input impedance, suppose that R_1 and R_2 are large. Then $i_{in} \simeq i_B$ and the input impedance of the transistor is

$$r_{in} = \frac{v_{in}}{i_{in}} = r_\pi + (\beta + 1)R'_E \simeq (\beta + 1)R'_E. \tag{11.17}$$

Since $\beta \simeq 100$, the input impedance can be quite large; for example, if $R'_E = 1\,\text{k}\Omega$, the transistor itself presents an input impedance of 100 kΩ. In this case, it is not correct to ignore R_1 and R_2, which appear in parallel with r_{in}. Let us also remark in passing that when R'_E is bypassed, as in the common emitter amplifier, $r_{in} = r_\pi$.

Output impedance

The output impedance of the amplifier is obtained from the Thevenin equivalent:

$$r_{out} = v_{out}(\text{open circuit})/i_{out}(\text{short circuit}).$$

On open circuit, from equation (11.16)

$$v_{out} = v_S R'_E/\{R'_E + (r_S + r_\pi)/(\beta + 1)\};$$

on short circuit

$$i_{out} = (\beta + 1)v_S/(r_S + r_\pi).$$

Copyright © 2005 IOP Publishing Ltd.

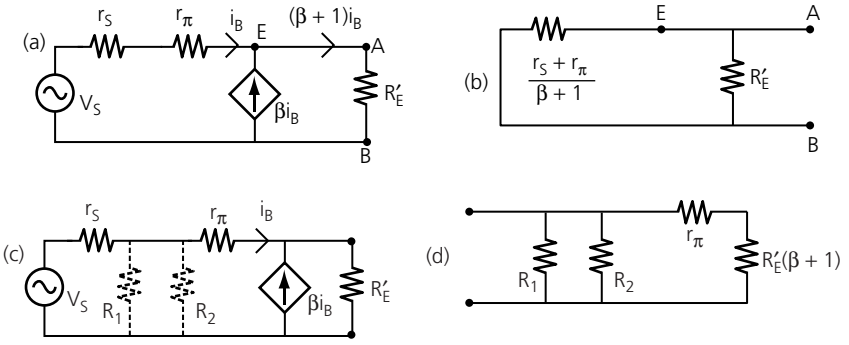


Fig. 11.18. (a) AC equivalent circuit omitting R_1 and R_2, (b) its equivalent for calculating r_{out}, (c) the effect of R_1 and R_2 on r_{out} and (d) r_{in}.

Here r_S is the output impedance of the source feeding the emitter follower; since $r_\pi \simeq 1\ \text{k}\Omega$, r_S may not be negligible. Finally,

$$r_{out} = \frac{R'_E(r_\pi + r_S)/(\beta + 1)}{R'_E + (r_\pi + r_S)/(\beta + 1)}. \tag{11.18}$$

If, for example, $R'_E = r_S = 1\ \text{k}\Omega$, and $\beta = 100$, then $R_{out} \simeq 20\ \Omega$.

The expression for r_{out} is just the resistance of R'_E in parallel with $(r_\pi + r_S)/(\beta + 1)$ and it is quite revealing to see why. In figure 11.18(a), the equivalent circuit is redrawn, omitting R_1 and R_2. The current through r_S and r_π is a factor $(\beta + 1)$ less than through R'_E, so the voltage across them is reduced by the same factor. As far as impedance is concerned, the same result is given by figure 11.18(b). The impedance across AB (i.e. the Thevenin equivalent) is given by R'_E in parallel with $(r_S + r_\pi)/(\beta + 1)$.

If for completeness we want to reinsert R_1 and R_2 into figure 11.18(a), as shown by the dashed lines in figure 11.18(c), it is clear that they appear in parallel with r_S. Figure 11.18(d) shows how to include the effect of R_1 and R_2 on input impedance.

The current gain is given by i_{out}/i_{in}. The maximum current gain is obtained by applying a short circuit across the output. In this case, $R'_E = 0$ and the input impedance becomes r_π, which is small compared with R_1 and R_2. Then the current gain is simply $(\beta + 1)$.

What is the maximum power gain?

Voltage feedback in the emitter follower

From figure 11.16, the AC voltage applied between base and emitter of the transistor is $v_{in} - v_{out}$. In the language of Chapter 7, this arrangement has 100% negative feedback, $B = 1$. This is directly responsible for the fact that

$$\frac{v_{out}}{v_{in}} \simeq 1 = 1/B$$

Copyright © 2005 IOP Publishing Ltd.

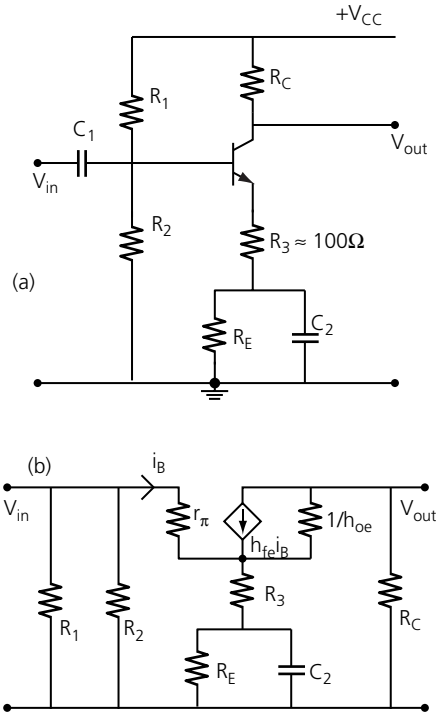


Fig. 11.19. (a) Common emitter amplifier with negative feedback from R_3 and (b) its equivalent circuit.

and also for the high input impedance and low output impedance. The signature of this type of feedback is the unbypassed emitter resistance R_E.

The choice of C_2

The equivalent circuit of figure 11.19(b) can be analysed with some labour. We shall not do this, but simply draw attention to the value required for C_2. Over the working frequency range of the amplifier, this should have an impedance small compared with $R_3 + r_\pi/(\beta + 1)$, otherwise its frequency dependence will distort the low frequency response. If $I_E = 2$ mA and frequency $\beta = 100$ for example, $r_\pi/(\beta + 1) \simeq 12.5\ \Omega$; if $R_3 = 0$ and the amplifier is to have a lower cut-off frequency of $\omega = 100\pi$ rad/s, a very large $C_2 \geq 250\ \mu$F is required. Even with an unbypassed $R_3 = 100\ \Omega$, $C_2 \simeq 30\ \mu$F is needed.

Summary

The emitter follower has high input impedance, low output impedance, voltage gain $\simeq 1$ and current gain up to $(\beta + 1)$. It serves as an impedance transformer

Copyright © 2005 IOP Publishing Ltd.

and is frequently used to drive a low impedance load such as a loudspeaker. Since the voltage gain is 1, there is no Miller effect and the high frequency performance is good.

11.8 FETs

With appropriate voltage supplies and bias resistors, the JFET can be inserted into all of the circuits discussed so far, making a common source amplifier (analogous to the common emitter) and a source follower (like the emitter follower). Its major practical difference from the bipolar transistor is its much higher input resistance. A second important difference is that the noise level of the FET is usually much lower than that of the bipolar transistor. Both features make the FET an obvious choice for the input stage of an operational amplifier. More generally, the first stage amplifier (preamplifier) for weak signals usually uses an FET, because of the low noise.

Biasing

Figure 11.20(a) shows a suitable arrangement for biasing the FET. The arrow on the transistor denotes the p-type gate (n-type channel). Suppose an operating point is required with $V_{GS} = -1$ V, $I_D = 2.5$ mA and $V_{DS} = 12$ V. The gate is at earth and the source at voltage $+R_S I_D$; hence we require $R_S \simeq 400\ \Omega$. The slope $1/(R_D + R_S)$ of the load line on figure 11.21 is given by $(R_D + R_S) = (20 - 12)/2.5 \times 10^{-3} \simeq 3.2$ kΩ, so $R_D \simeq 2.8$ kΩ. The purpose of R_1 is to prevent the gate from floating. Then C_1 and R_1 define the low frequency cut-off.

If the input signal is very small, can you scale up R_S and R_D by a factor of say 30 and avail yourself of the full voltage gain available from R_D?

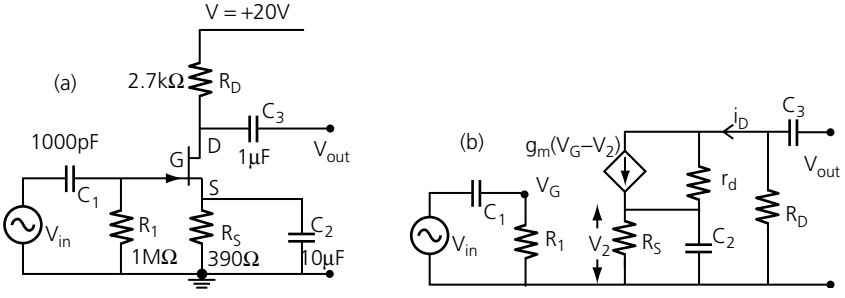


Fig. 11.20. (a) the common source amplifier and (b) its equivalent circuit.

Copyright © 2005 IOP Publishing Ltd.

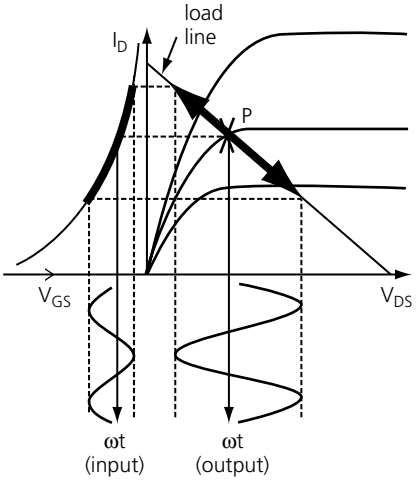


Fig. 11.21. Amplification in the FET.

Common-source amplifier

The AC equivalent circuit of the resulting common-source amplifier is shown in figure 11.20(b). Providing C_2 is large enough to bypass R_S, the output stage is simple. To see how large C_2 has to be:

$$i_D = g_m(v_G - i_D Z_2)$$

where Z_2 is the impedance of R_S and C_2 in parallel. So

$$\boxed{i_D = \frac{g_m v_G}{1 + g_m Z_2};} \tag{11.19}$$

neglect of Z_2 requires $Z_2 \ll 1/g_m \simeq 300\ \Omega$. If $\omega = 100\pi$ rad/s, $C_2 \geq 10\ \mu$F. Then the output impedance r_{out} is given by R_D and r_d in parallel; if the load applied across the output is R_L, the voltage gain is $g_m R'_L$, where R'_L is the parallel combination of R_L with r_{out}. The input impedance is R_1.

Source follower

This analysis also reveals how to create a source follower (the FET analogue of the emitter follower), shown in figure 11.22. The requirement of negative feedback for source follower operation follows from equation (11.4), namely $R_S \gg 300\ \Omega$, where $R_S = R_f + R_2$. Then voltage gain is

$$G_V = \frac{g_m R_S}{1 + g_m R_S} = \frac{1}{1 + 1/g_m R_S}.$$

Copyright © 2005 IOP Publishing Ltd.

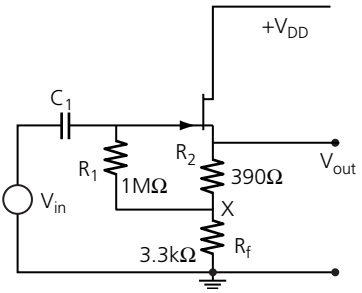


Fig. 11.22. The source follower.

The output impedance is

$$R_{out} = \frac{V_{out}\,(\text{open circuit})}{I_{out}\,(\text{short circuit})}$$

$$= \frac{R_S g_m v_G/(1+g_m R_S)}{g_m v_G} \qquad \text{using equation (11.19)}$$

$$= \frac{R_S}{1+g_m R_S} = \frac{1}{g_m + 1/R_S}. \tag{11.20}$$

This is equal to the parallel combination of R_S with a resistance $1/g_m$. If R_S is the larger of the two, $R_{out} \simeq 1/g_m \simeq 300\ \Omega$. This is larger than the output impedance of the emitter follower (typically 20 Ω), so the latter is normally preferred when a high output current is required. If the source follower is used to drive a very low resistance load ($R_S \leq 1/g_m$), the voltage gain $(1+1/g_m R_S)^{-1}$ drops significantly below 1.

It is recommended that you measure for yourself the characteristics of an FET and use it to construct a common source amplifier and source follower (by removing R_D); check the DC levels in the circuit with an oscilloscope and then follow through the amplification of an AC voltage. You can obtain r_d either from the characteristic curves or by measuring the output impedance using a tuned circuit on resonance as load (very high AC impedance, low DC impedance).

11.9 Exercises

1. A transistor in the common emitter configuration has $h_{ie} = 2$ kΩ, $h_{fe} = 120$, $h_{re} = 2 \times 10^{-4}$ and $1/h_{oe} = 15$ kΩ. It is fed by an AC signal $V_{in} = 3\cos\omega t$ mV and drives a load resistor $R_L = 2.2$ kΩ. First neglect h_{re} and then include it as a correction. Find the AC base current i_B, the AC collector current i_C and the AC current through the load resistor. What are the current and voltage gains through R_L? What are the input and output impedances of the circuit? (Ans: with $h_{re} = 0$, $i_B = 1.5\ \mu$A,

Copyright © 2005 IOP Publishing Ltd.

$i_C = 0.18\,\text{mA}, i_L = 0.157\,\text{mA}, v_{CE} = -0.345\,\text{V}, h_{re}v_{CE} = -0.069\,\text{mV}$; including h_{re}, $i_B = 1.53\ \mu\text{A}$, $i_C = 0.184$ mA, $i_L = 0.161$ mA; current gain = -105, voltage gain = -118, $r_{in} = 1.96\ \text{k}\Omega$, $r_{out} = 1.92\ \text{k}\Omega$.)

2. A transistor operating with $I_E = 2.5$ mA has $r_b = 100\ \Omega$, $\beta = 210$ and $r_0 = \infty$. What is r_π? The transistor is used in a common emitter amplifier fed by an AC voltage $5\cos\omega t$ mV, and drives a 1.8 kΩ load. Draw the equivalent circuit and find (a) i_B, (b) the output voltage across the load. What is the input impedance of the amplifier, the current gain and the voltage gain? (Ans: 2.1 kΩ, (a) $2.26\cos\omega t\ \mu$A, (b) $-0.85\cos\omega t$ V; 2.2 kΩ, -210, -171.)

3. A transistor operating in the common emitter configuration with current gain $\beta = 200$ has an input impedance of 4 kΩ. Draw the hybrid-π equivalent circuit. Set $r_0 = \infty$ initially. If the input voltage is a 4 mV AC signal, what is the output voltage across a load resistor of 2.2 kΩ? Using these values, what would be the current through an actual r_0 of 20 kΩ? ∗With this value of r_0, what is the true value of the voltage cross the load resistor? (Ans: -440 mV, $19.8i_{in}$, -396 mV.)

4. The circuit of figure 11.23 acts as an approximately constant source of current, insensitive to the magnitude of R_L. The voltage across the transistor is $V_{DS} = V_{DD} - (R_L + R_1)I_D$. By considering small changes in V_{GS}, V_{DS} and I_D, show that $\delta I_D = -I_D\delta R_L/\{R_L + R_1 + r_d(1 + g_m R_1)\}$. If $R_L = 3.3$ kΩ, $R_1 = 390\ \Omega$, $g_m = 3 \times 10^{-3}$ mho and $r_d = 40$ kΩ, what is the fractional change in I_D if R_L is halved? (Ans: $+0.018$.)

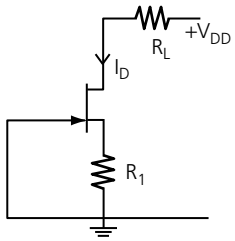


Fig. 11.23.

5. What values do R_1, R_2, R_C and R_E take in the circuit of figure 11.24 if the operating point is to be at $V_{CE} = 8$ V, $V_{BE} = 0.6$ V, $I_C = 5$ mA, $I_B = 100\ \mu$A and the base is to be at 5 V above earth? (b) You may assume that $h_{fe} = 50$, $1/h_{oe} = 12$ kΩ, $h_{ie} = 2$ kΩ, and $h_{re} = 0$. Ignoring capacitances, but not the bias and load resistors, what are the input and output impedances of the amplifier and the maximum current and voltage gains available when feeding suitable loads? (c) What is the magnitude of C_1 which will reduce the voltage gain by 3 db at a frequency $f = 50$ Hz? ∗(d) What change in operating point is there if R_2 is 10%

Copyright © 2005 IOP Publishing Ltd.

larger than it should be? Hint: Remember to make $I_B = I_C/h_{fe}$. (Ans: (a) $R_1 = 30$ kΩ, $R_2 = 12.5$ kΩ, $R_E = 860$ Ω, $R_C = 1.5$ kΩ; (b) $r_{in} = 1.6$ kΩ, $r_{out} = 1.35$ kΩ, maximum current and voltage gains 40 and 34, (c) $C_1 = 2$ μF, (d) V_B and V_E increased by 0.28 V, I_E and I_C raised by 8.6%, V_{CE} reduced by 0.72 V.)

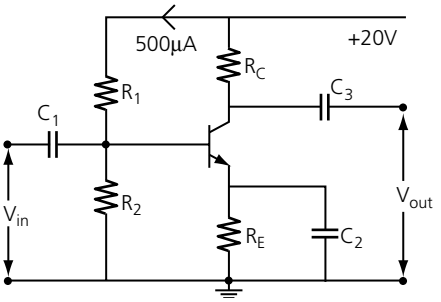


Fig. 11.24.

6. A transistor with $h_{ie} = 1$ kΩ, $h_{re} = 5 \times 10^{-4}$, $h_{fe} = 50$ and $1/h_{oe} = 2 \times 10^4$ Ω is used in the common emitter amplifier of figure 11.25. Draw the h equivalent circuit assuming all the capacitors to have negligible impedance, and also initially neglecting $h_{re}v_{CE}$. Calculate the input and output impedances and the maximum available current and voltage gains of the amplifier. Now evaluate the magnitude of $h_{re}v_{CE}$. What is its effect on i_B? For the rest of this question neglect h_{re}. If the output of this amplifier feeds into an identical second stage amplifier, what are the overall current and voltage gains of the two stages? What values must all the capacitances have for the first amplifier if they are individually to affect the gains by less than 3 db at $f = 50$ Hz? (Ans: $r_{in} = 870$ Ω, $r_{out} = 1.8$ kΩ, 43.5, 91, $h_{re}v_{CE} = -0.046v_{in}$; i_B increases by ∼ 4.6%; 1110, 2020; $C_1 = 3.6$ μF, $C_2 = 5.4$ μF, $C_3 = 160$ μF.)

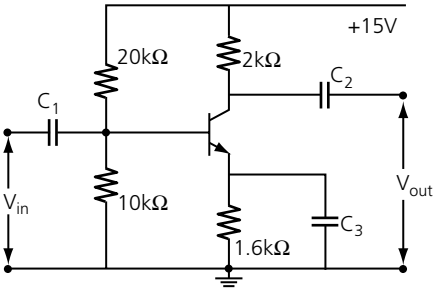


Fig. 11.25.

Copyright © 2005 IOP Publishing Ltd.

7. Identify one part of figure 11.26(b) for which figure 11.26(a) is the AC equivalent circuit. In particular, describe the meaning of v and R_1 in this context and give the values of R_2 and C. (Ans: Output side of the transistor: $C = 10\ \mu$F; R_2, R_1 and V are the Thevenin equivalent of $h_{fe}i_B$ in parallel with h_{oe}; $R_2 = 6\text{k}\Omega$.)

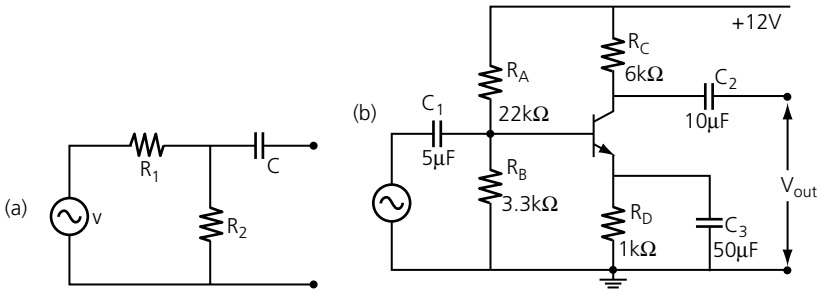


Fig. 11.26.

8. (QMC). In figure 11.27, what sort of feedback, for small signals, is provided by R_E? Draw the small signal equivalent circuit of the amplifier and, assuming that h_{re} is zero and that $1/h_{oe}$ is infinite, determine an expression for the closed loop voltage gain. If the amplifier is required to have a closed loop voltage gain of -10 with $h_{ie} = 5\ \text{k}\Omega$ and $h_{fe} = 250$, calculate a value for R_E. (Ans: gain $= -2.03\ \text{k}\Omega \times \beta/\{h_{ie} + (\beta + 1)R_E\}$; $R_E = 182\ \Omega$.)

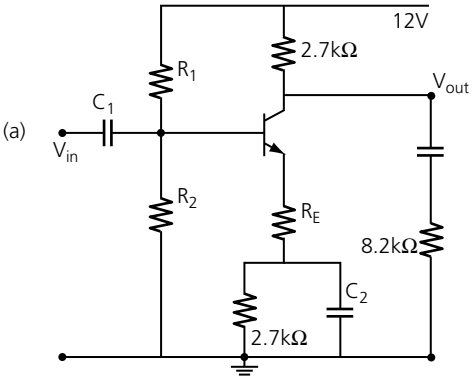


Fig. 11.27.

9**. The transistor in figure 11.28 has $h_{ic} = 1000\ \Omega$, $h_{fc} = 50$, $h_{oc} = 25 \times 10^{-6}$ mho and $h_{rc} = 0$. If the operating point is to be $I_E = 2.5$ mA and $I_B = 50\ \mu$A, what should be the value of R_2? What are the input and output impedances of this amplifier, and the maximum available current and voltage gains, (a) with $h_{oc} = 0$, (b) with h_{oc} given above? (Ans:

Copyright © 2005 IOP Publishing Ltd.

$R_2 = 125$ kΩ; (a) 35.2 kΩ, 19.69 Ω, 50.2, 0.988, (b) 34.6 kΩ, 19.68 Ω, 50.2, 0.987.)

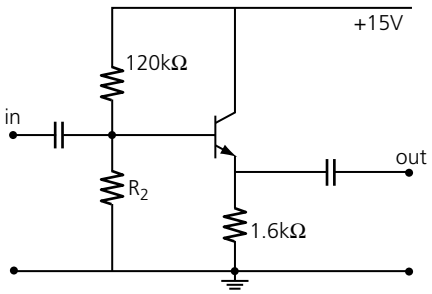


Fig. 11.28.

10. The amplifier in figure 11.29 in called a Darlington pair. It consists of two emitter followers in sequence. Find R_1. If both transistors have $h_{ic} = 1\ k\Omega$, $h_{fc} = 50$, $h_{oc} = 0$ and $h_{rc} = 0$, find the maximum possible voltage and current gains with suitable load resistors R_L, hence the input and output impedances. With $R_L = 12\Omega$, what is the power gain. (Ans: $R_1 = 2.2$ MΩ, 2653, 0.962, $r_{in} = 852$ kΩ, $r_{out} = 9.8\Omega$, 2.0×10^4.)

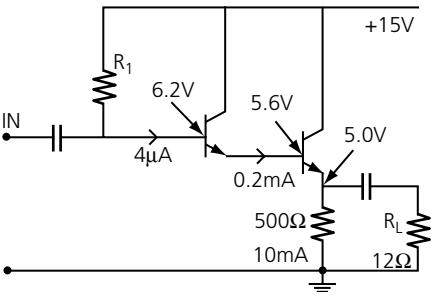


Fig. 11.29.

11**. (Lengthy but valuable.) Describe qualitatively, then quantitatively, the operation of the circuit of figure 11.30. Assume (i) both transistors have a current gain of 50, (ii) T_1 has an input impedance of 700 Ω, and an output impedance of 20 kΩ, (iii) for T_2 the value of r_b is 500 Ω and $r_{out} = \infty$. (a) What is the action of R_1 and R_2? (b) Find the operating point by writing 3 equations for V_{B1} (the voltage at the base of T1), I_{E1} and I_{E2}, and show that $V_{B1} \simeq 5.4$ V. Assume a drop of 0.6 V between base and emitter of each transistor. Find I_2, hence r_π for T2. (c) Ignoring the effect of R_2, find the input impedance of T2. Hence find the voltage gains across T1 and across T2. (d) Considering now the effect of R_2, what is the relation between v_{in} and v_{out} for AC signals? (e) What are the input and output

Copyright © 2005 IOP Publishing Ltd.

impedances? (Ans: (a) R_1 and R_2 form a current feedback loop with $B = 1/10$ to stabilise the overall voltage gain at $\sim R_2/R_1 = -10$; (c) 51.6 kΩ; voltage gains 74 and 1; (d) $v_{out}/v_{in} = -8.8$; (e) $r_{in} \simeq 4$ kΩ, $r_{out} \simeq 125$ Ω.)

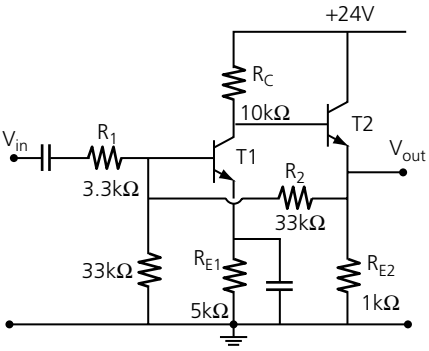


Fig. 11.30.

12. (RHBNC). The circuit of figure 11.31 shows a typical common source amplifier circuit. Explain the function of (a) the resistors R_G and R_S, (b) the capacitors C_G and C_D, (c) the capacitor C_S. The transfer and output characteristics of the transistor are shown in figure 11.32. Draw a load line for $V_{DD} = 30$ V, $R_D = 3.0$ kΩ and use it to explain how this amplifier operates when a small sinusoidal voltage is applied to its input. If the gate-source voltage is -1.0 V, calculate the value of the resistor R_S which is required to provide the gate-source bias voltage. With the circuit parameters given above, find (i) the DC voltage at the drain with respect to earth when no signal is applied to the input, (ii) the voltage gain of the amplifier when a sinusoidal input signal of amplitude 0.5 V is applied to its input. (Ans: $R_S = 250$ Ω, (i) 18 V, (ii) 18.6.)

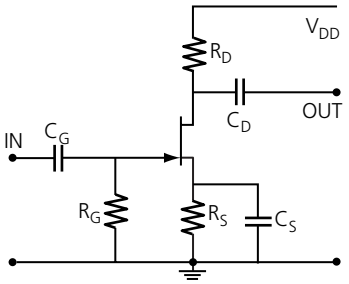


Fig. 11.31.

13. A common source amplifier has a load resistor of 22 kΩ and gives a voltage amplification of 42.5. When the load resistor is halved, the gain

Copyright © 2005 IOP Publishing Ltd.

drops to 26. Find the output impedance of the FET and its g_m. (Ans: $r_d = 38$ kΩ, $g_m = 3.05$ mA/V.)

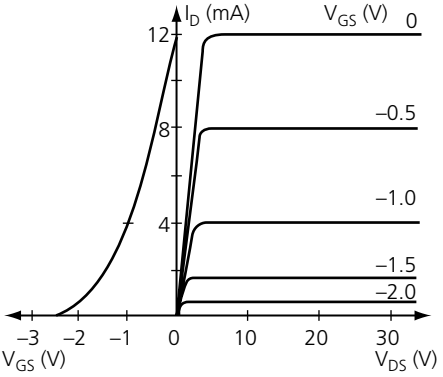


Fig. 11.32.

14. Show that for the source follower of figure 11.22,

$$i_{in}R_1 = R_2 i_D + v_{GS} \qquad v_{DS} + (R_2 + R_f)i_D \simeq 0$$

$$i_D = g_m v_{GS} + v_{DS}/r_d \qquad \text{and} \qquad v_{in} \simeq v_{GS} + (R_2 + R_f)i_D$$

where small i and v refer to small changes in currents and voltages. (The approximations involve neglecting the voltage drop across R_f due to i_{in}.) Hence show that the input impedance is $r_{in} = R_1\{1 + g_m(R_2 + R_f) + (R_2 + R_f)/r_d\}/\{1 + g_m R_2 + (R_2 + R_f)/r_d\}$. Evaluate this for the parameter values of figure 11.22 if $g_m = 1/300$ mho and $r_d = 20$ kΩ. (Ans: 1.1 MΩ.)

Copyright © 2005 IOP Publishing Ltd.

12

Gates

12.1 Introduction

Digital logic is used in computers and a huge range of technical applications. It is all based on electronic circuits like the OR and AND gates we met in Chapter 9. Transistors make natural high-speed switches and everyday PC's use circuits which switch currents at rates of typically 10^9 Hz today. Here, we will explore the basic elements and logic used in digital systems.

Numbers are expressed in **binary** form. For instance, the decimal number 13 (written 13_{10}, where the subscript denotes the base) is represented in binary by

$$13_{10} = 1 \times 2^3 + 1 \times 2^2 + 0 \times 2^1 + 1 \times 2^0 \equiv 1101_2.$$

Each binary digit or **bit** can be held by an electronic element which is on or off.

We are so accustomed to decimal that some pieces of hardware (for example digital clocks, calculators and digital voltmeters) use a mixed binary and decimal system called **binary coded decimal** (BCD). Scalers may be designed to count from 0 to 9 in binary, then carry to a new unit which counts the next decade 10 to 90 and so on. The electronic logic required at the transition 9 to 10 is not complicated.

It is arbitrary whether a binary 1 is represented by the more positive or more negative voltage (or by a current which is on or off). We shall use for illustration voltage levels +5 V and 0 V and positive logic. Then logical 1 is represented by +5 V and logical 0 by 0 V, figure 12.1(a). If any binary digit is denoted by A, its complement, written $\overline{A}$ (or sometimes A') is its converse; if $A = 0$, then $\overline{A} = 1$ and is represented here by +5 V; if $A = 1$, $\overline{A} = 0$ and is represented by 0 V. A circuit which converts A into $\overline{A}$ is said to execute the NOT operation: $\overline{A} = NOT(A)$. Electronically, it is described as **inversion** of A; this does not mean that the 5 V level representing 1 is inverted to −5 V, but that it is logically inverted to 0 V. The circuit symbol is shown in figure 12.1(b); the bubble on the output denotes the complement (i.e. logical inversion).

Copyright © 2005 IOP Publishing Ltd.

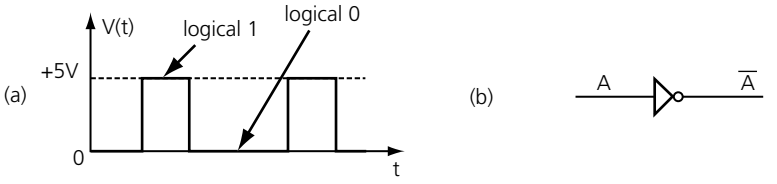


Fig. 12.1. (a) Positive logic, (b) the circuit representation of an inverter.

12.2 Logical Combinations of A and B

AND

Suppose two binary digits A and B are to be multiplied together. Either can be 0 or 1, so there are four possible combinations giving the results shown in the table of figure 12.2(a). This is called a **truth table**. The operation of binary multiplication is called AND; it is written $A.B$ (or often just AB). The result is 1 only if A and B are both 1. Electrically, it may be realised with the circuit of figure 12.2(b): current flows only if both switches A and B are closed. Electronically, the switches are replaced by transistors which may be switched in a few ns (10^{-9} s). Readers may follow this chapter and the next two without needing to know how transistors function; however figure 9.31 shows a rudimentary circuit performing the AND function and exercise 4 of this chapter discusses it.

NAND

A unit which performs the AND function is represented in circuit diagrams by the symbol shown in figure 12.2(a). Its complement, shown in figure 12.2(c) is called NAND (short for NOT AND). Its circuit symbol has a bubble on the gate output to

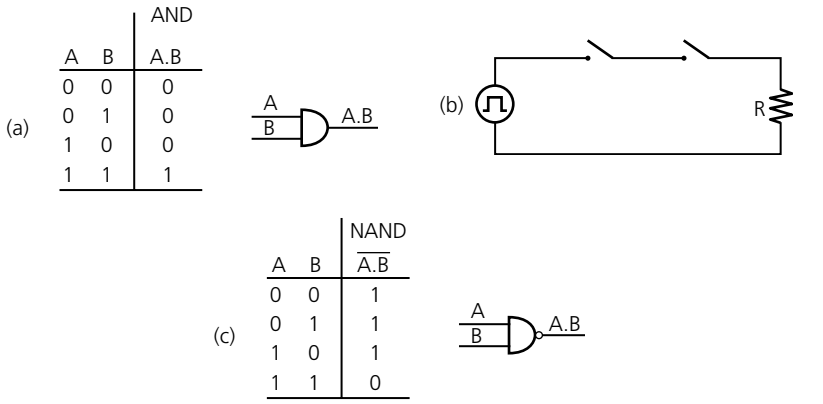


Fig. 12.2. (a) Truth table and circuit symbol for AND, (b) forming AND with switches, (c) truth table and ciruit symbol for NAND.

Copyright © 2005 IOP Publishing Ltd.

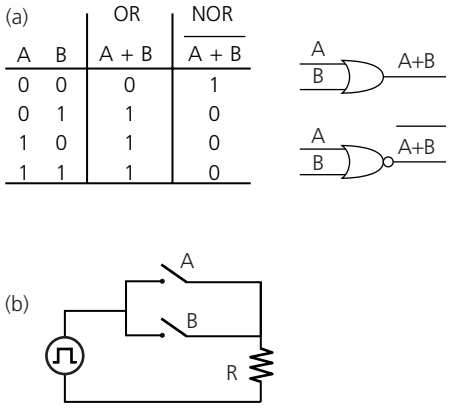


Fig. 12.3. (a) Truth tables and circuit symbol for OR and NOR, (b) the electrical principle of the OR gate.

indicate the complement. Electronically, the switching of NAND circuits is more decisive than AND, so commercial units are based upon NAND circuitry. If AND is required, it may be formed by NAND followed by inversion to get back to AND.

OR and NOR

A second basic logical operation is called OR. Its truth table and circuit symbol are given in figure 12.3(a). It is denoted algebraically by $A + B$ and the result is 1 if either A or B (or both) is 1. The $+$ symbol should not be confused with arithmetic addition. Electrically, the OR operation may be realised with switches by the circuit of figure 12.3(b); current flows through the load R if *either* switch A *or* B is closed. Electronically, the switches are replaced by transistors. A circuit doing this is given in Figure 9.30, but again it is not essential to understand the transistor operation in order to use commercial units. The complement of OR is called NOR and is written $\overline{A + B}$. Its truth table and circuit symbol are given in figure 12.3(a).

Circuits performing any of these logical operations are called **gates**. The origin of this name can be understood from figure 12.4(a). There, a long signal A 'opens a gate' to allow the pulse train from B through to the output by forming $A.B$. If $\overline{A}$ is substituted for A to make $\overline{A}.B$, as in figure 12.4(b), the pulse train is halted when $A = 1$, $\overline{A} = 0$. Names given to this operation are **veto**, **inhibit** and **busy**; these names should be self-explanatory.

Half-Adder

From the operations AND, OR and NOT, every operation required in digital logic can be constructed. Indeed, technically they can all be done with NAND gates only or with NOR gates only, though the contortions required to achieve this may be

Copyright © 2005 IOP Publishing Ltd.

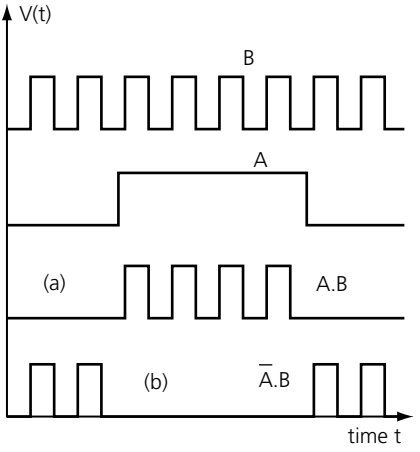


Fig. 12.4. (a) $A.B$ and $\overline{A}.B$.

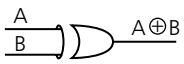
inconvenient. As an example of the use of gates, figure 12.5(a) gives the truth table for the arithmetic addition of two binary digits A and B. This is called a **half adder**. The last line of figure 12.5(a) expresses the fact that decimal 1 plus decimal 1 = decimal 2 ≡ binary 10. The carry is simply $A.B$ and may be formed with an AND gate. The sum, called **exclusive or** (XOR) and written $A \oplus B$, may be expressed as $A.\overline{B} + B.\overline{A}$; it is 1 if $A = 1$ and $B = 0$ or if $B = 1$ and $A = 0$, and is 0 otherwise.

A circuit which functions as a half adder is shown in figure 12.6 and the way the logic works is illustrated in figure 12.7. There, the four logical possibilities for A and B are shown by the two waveforms at the top. The next two lines show $A + B$ and $\overline{A.B}$, which are the outputs of the left-hand gates in figure 12.6. The last line of figure 12.7 shows the AND of lines 3 and 4. This forms $A \oplus B$. You may like to check the operation of figure 12.7 by writing down the truth table for every gate in that figure.

Care is needed over timing. Figure 12.8 shows a pair of signals A and B intended to form $A.B$. However, B is a little late. This will narrow the pulse intended to be $\overline{A.B}$. The way to overcome this problem is to use a 'clock' pulse C, dedicated

(a)

A	B	SUM	CARRY
0	0	0	0
0	1	1	0
1	0	1	0
1	1	0	1



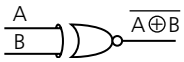


Fig. 12.5. (a) Truth table for a half adder, (b) and (c) circuit symbols for XOR and XNOR.

Copyright © 2005 IOP Publishing Ltd.

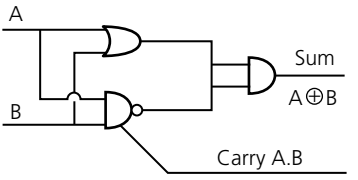


Fig. 12.6. The half adder.

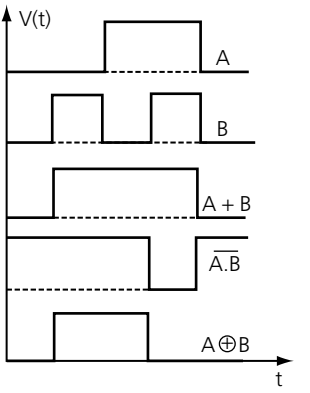


Fig. 12.7. Waveforms of the half adder.

to timing. The output $A.B.C$ is used to regenerate a standard output pulse timed correctly by the clock. Such a system is called **synchronous**.

12.3 Boolean Algebra

So far, we have managed nicely with truth tables and simple logical operations OR and AND. However, when it comes to the design of a complicated system, algebraic and graphical methods are needed to help sort out the logic. Just such an algebra was developed by George Boole, who in 1854 published a treatise entitled 'An Investigation of the Laws of Thought'. His algebra is actually more general than is needed for binary systems, but is readily simplified to present needs.

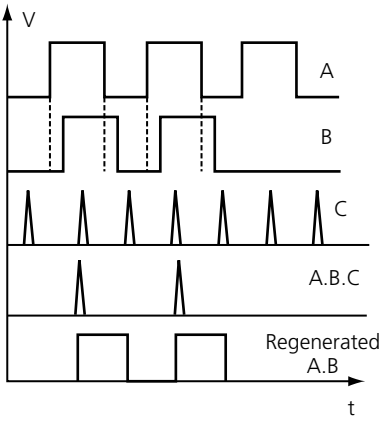


Fig. 12.8. A synchronous system.

Copyright © 2005 IOP Publishing Ltd.

(i) $A + B = B + A$, $A.B = B.A$ } commutivity: the order is irrelevant

(ii) $A + 0 = A$, $A.1 = A$

(iii) $A + 1 = 1$, $A.0 = 0$

(iv) $A + \overline{A} = 1$, $A.\overline{A} = 0$

(v) $\overline{\overline{A}} = A$

(vi) $A + A.B = A$, $A.(A + B) = A$ } absorption - very useful

(vii) $\overline{A + B} = \overline{A}.\overline{B}$, $\overline{A.B} = \overline{A} + \overline{B}$ } de Morgan's theorems

(viii) $A + (B + C) = (A + B) + C$, $A.(B.C) = (A.B).C$ } order of operations irrelevant

(ix) $A.(B+C) = A.B + A.C$, $A + (BC) = (A + B).(A + C)$ } distribution: how to evaluate expressions

Fig. 12.9. Rules of Boolean algebra.

Suppose A, B, C, etc. are symbols denoting binary digits, i.e. each of them is either 0 or 1. Rules of Boolean algebra are given in figure 12.9. From truth tables, they are all obvious. In using them, the order in which Boolean expressions must be evaluated is as follows: parentheses first, then NOT, AND, and finally OR. Two statements are identical if they have the same truth table. The second of relations (ix) differs from the familiar result for decimal numbers, but is verified in figure 12.10(a).

Try relations (vi) for yourself along the same lines.

12.4 De Morgan's Theorems

Relations (vii) give two famous theorems due to de Morgan; these are demonstrated in the truth tables of figure 12.10(b). They are much easier to remember in words. They imply that:

(a) NOR ≡ AND between complementary signals,

(b) NAND ≡ OR between complementary signals.

Later a simple graphical way of expressing them will be given. These relations demonstrate a general property of all of relations (i) to (ix) known as **duality**. Suppose AND is changed to OR and vice versa in one relation of a pair, and suppose 0 is simultaneously changed to 1; the second relation of the pair is then obtained.

Copyright © 2005 IOP Publishing Ltd.

(a)

A	0	0	0	0	1	1	1	1
B	0	0	1	1	0	0	1	1
C	0	1	0	1	0	1	0	1
B.C	0	0	0	1	0	0	0	1
A + (B.C)	0	0	0	1	1	1	1	1
A + B	0	0	1	1	1	1	1	1
A + C	0	1	0	1	1	1	1	1
(A + B).(A + C)	0	0	0	1	1	1	1	1

(b)

	A	0	0	1	1
	B	0	1	0	1
	A + B	0	1	1	1
(a)	$\overline{A + B}$	1	0	0	0
	$\overline{A}$	1	1	0	0
	$\overline{B}$	1	0	1	0
(a)	$\overline{A}.\overline{B}$	1	0	0	0
	A.B	0	0	0	1
(b)	$\overline{A.B}$	1	1	1	0
(b)	$\overline{A} + \overline{B}$	1	1	1	0

Fig. 12.10. Verification that A + (B.C) = (A + B).(A + C), (b) proof of de Morgan's theorems.

It is always possible to express any logical expression in terms of NOT, AND and OR, often in several different ways. Usually we proceed by constructing a truth table, then simplifying the result using algebraic, or more often graphical methods. Some examples will indicate the process.

12.5 The Full Adder

Let us consider how to add two binary numbers together. For any one digit of the result, it is necessary to add two input binary digits A and B and a carry C coming from the previous digit. Results are given in figure 12.11 for all possible values of A, B and C. Algebraically, the ones in this table may be denoted by

$$\text{Sum} = \overline{A}.\overline{B}.C + \overline{A}.B.\overline{C} + A.\overline{B}.\overline{C} + A.B.C \tag{12.1}$$

$$\text{Carry} = \overline{A}.B.C + A.\overline{B}.C + A.B.\overline{C} + A.B.C. \tag{12.2}$$

A circuit forming these logical expressions with OR and AND gates gives an output 1 wherever there is a 1 in figure 12.11; otherwise the outputs are zero. The logic

Copyright © 2005 IOP Publishing Ltd.

A	0	0	0	0	1	1	1	1
B	0	0	1	1	0	0	1	1
C	0	1	0	1	0	1	0	1
Sum	0	1	1	0	1	0	0	1
Carry	0	0	0	1	0	1	1	1

Fig. 12.11. The truth table of a full adder.

therefore reproduces the truth table. These expressions are said to be in AND-OR form, since they are formed as ANDs followed by ORs.

An obvious question, however, is whether these expressions can be simplified. In the case of equation (12.2) this is so. It may be rewritten:

$$\text{Carry} = A.B.(C + \overline{C}) + B.C.(A + \overline{A}) + C.A.(B + \overline{B}). \qquad (12.2a)$$

The fact that some terms appear twice in (12.2a) and once in (12.2) is of no consequence: $A.B.C$ is the same as $B.C.A$ and $C.A.B$ (the order is irrelevant) and $A.B.C + A.B.C = A.B.C$ from Boolean relation (vi). Then equation (12.2a) gives

$$\text{Carry} = A.B + B.C + C.A. \qquad (12.3)$$

You may verify this relation using a truth table.

12.6 The Karnaugh Map

An important graphical method of spotting such simplifications is the Karnaugh map, shown in figure 12.12. This diagram displays the 8 possible values (reading top left to bottom right):

$$\overline{A}.\overline{B}.\overline{C},\ \overline{A}.\overline{B}.C,\ \overline{A}.B.C,\ \overline{A}.B.\overline{C},\ A.\overline{B}.\overline{C},\ A.\overline{B}.C,\ A.B.C,\ A.B.\overline{C}.$$

Entries for BC are arranged so that only one bit changes between 0 and 1 from one entry to the next. Entries are then marked into the map corresponding to terms in the logical expression for Carry, equation (12.2). If two entries appear on

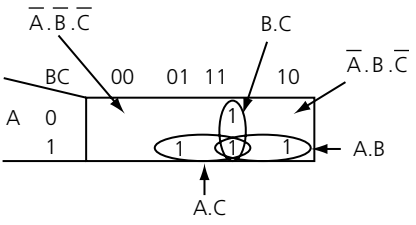


Fig. 12.12. Karnaugh map for the carry signal of a full adder.

Copyright © 2005 IOP Publishing Ltd.

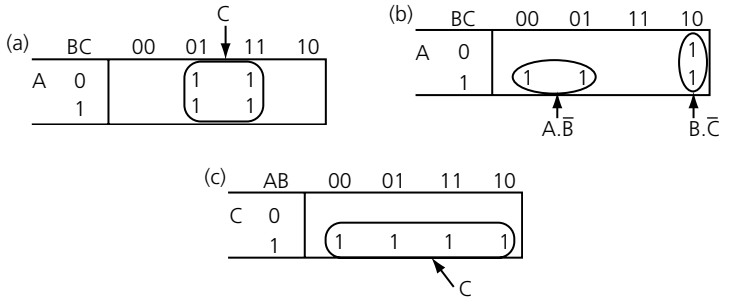


Fig. 12.13. Simplification of Karnaugh maps.

neighbouring sites, they can be simplified by relations $A+\overline{A}=1$ or $B+\overline{B}=1$ or $C+\overline{C}=1$. These simplifications are indicated in figure 12.12 by the loops drawn round neighbouring pairs, and we read off the result given by equation (12.3): e.g. the bottom left loop gives $A.C.(B+\overline{B})=A.C$.

As a second example, figure 12.13(a) shows the entries for $\overline{A}.\overline{B}.C+\overline{A}.B.C+A.\overline{B}.C+A.B.C$. In this case, two sets of neighbouring pairs may be combined using $A+\overline{A}=1$ and $B+\overline{B}=1$; C is one for all entries and is combined with all pairs of values of A and B. So the result is simply C.

Next consider the reverse procedure. Suppose the Karnaugh map is to be drawn for the expression $A.\overline{B}+B.\overline{C}$. How is this done? Firstly, $A.\overline{B}=A.\overline{B}.(C+\overline{C})$ and is represented by the two entries at the bottom left of figure 12.13(b); $A.\overline{B}$ is common to both. Likewise $B.\overline{C}$ is given by the entries at the right of the table, summing over the possible values $A=0$ or 1.

It does not matter whether horizontal rows of the Karnaugh map display values of B and C or AB or even AC. Figure 12.13(c) shows the Karnaugh map of Figure 12.13(a) replotted with AB horizontal and C vertical. The map looks different to figure 12.13(a). However, after the simplifications have been made, combining neighbouring terms, the result is the same.

It is recommended that you get plenty of practice in manipulating such diagrams. They are almost invariably easier to use than Boolean algebra itself, though once an expression has been simplified using the Karnaugh map, it is good practice to check that the same result can be obtained algebraically, lest a mistake has crept in somewhere.

As a further example, figure 12.14 gives the Karnaugh map for the sum signal of the full adder, equation (12.1). Because no entries are adjacent, the expression

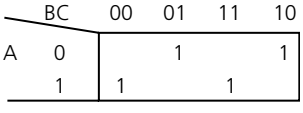


Fig. 12.14. Karnaugh map for the sum signal of a full adder.

Copyright © 2005 IOP Publishing Ltd.

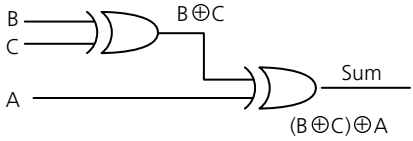


Fig. 12.15. Forming the sum from two half adders.

cannot be simplified. Even so, the pattern of ones gives a valuable clue to a way of forming the required sum signal. The ones occupy alternate horizontal locations. This corresponds to the fact that the arithmetic sum of B and C is given by the exclusive OR, $B \oplus C$. The staggering of the ones between the two horizontal lines suggests that the required overall sum signal is given by $A \oplus (B \oplus C)$, and this is indeed so, as you can readily verify with a truth table. The Sum signal can therefore be constructed from two half-adders, as shown in figure 12.15. The order in which the two exclusive ORs is carried out does not matter; $A \oplus (B \oplus C)$ or $(A \oplus C) \oplus B$ or $(A \oplus B) \oplus C$ give the same result.

The Karnaugh map is readily extended to expressions involving four characters A, B, C and D. As an illustration, suppose two 2-digit binary numbers are to be multiplied. The required results are given in Table 12.1. Let AB

Table 12.1 Multiplication of two binary two-digit numbers.

$00 \times 00 = 0000$	$00 \times 01 = 0000$	$00 \times 10 = 0000$	$00 \times 11 = 0000$
$01 \times 00 = 0000$	$01 \times 01 = 0001$	$01 \times 10 = 0010$	$01 \times 11 = 0011$
$10 \times 00 = 0000$	$10 \times 01 = 0010$	$10 \times 10 = 0100$	$10 \times 11 = 0110$
$11 \times 00 = 0000$	$11 \times 01 = 0011$	$11 \times 10 = 0110$	$11 \times 11 = 1001$

and CD represent the two binary numbers and $WXYZ$ the binary digits of the result. Then the ones of Table 12.1 give

$$W = A.B.C.D$$

$$X = A.\overline{B}.C.\overline{D} + A.\overline{B}.C.D + A.B.C.\overline{D} \tag{12.4}$$

$$Y = \overline{A}.B.C.\overline{D} + \overline{A}.B.C.D + A.\overline{B}.\overline{C}.D + A.\overline{B}.C.D + A.B.\overline{C}.D + A.B.C.\overline{D} \tag{12.5}$$

$$Z = \overline{A}.B.\overline{C}.D + \overline{A}.B.C.D + A.B.\overline{C}.D + A.B.C.D. \tag{12.6}$$

The Karnaugh maps for X, Y and Z are shown in figure 12.16. The simplifications are

$$Z = B.D$$

$$Y = A.\overline{C}.D + A.\overline{B}.D + \overline{A}.B.C + B.C.\overline{D} = A.D.(\overline{C} + \overline{B}) + B.C.(\overline{A} + \overline{D})$$

$$X = A.\overline{B}.C + A.C.\overline{D} = A.C.(\overline{B} + \overline{D}).$$

Copyright © 2005 IOP Publishing Ltd.

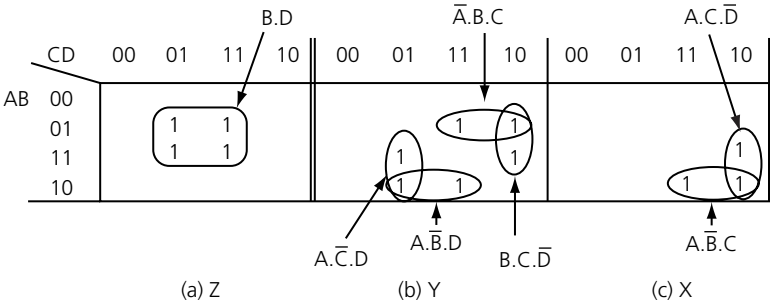


Fig. 12.16. Karnaugh maps for (a) equation (12.6), (b) (12.5) and (12.4).

as can be verified algebraically from equations (12.4) – (12.6). It is a straightforward matter to arrange NOT, AND and OR gates so as to create these signals.

In order to consider a further point, figure 12.17(a) shows the expression

$$f = A.B.\overline{C}.\overline{D} + \overline{A}.\overline{B}.C.D + A.\overline{B}.C.D + A.B.C.\overline{D}. \tag{12.7}$$

Algebraically, the first and fourth terms reduce to $A.B.\overline{D}$ and the second and third to $\overline{B}.C.D$. To get these results from the Karnaugh map, we must imagine it wrapped around a horizontal axis so that top and bottom edges join; then the entries $\overline{A}.\overline{B}.C.D$ at the top and $A.\overline{B}.C.D$ at the bottom are adjacent and may be combined to $\overline{B}.C.D$ using $A + \overline{A} = 1$, as shown by the semicircles on the map. Likewise, we must imagine the map wrapped around a vertical axis, so that left and right hand edges join; this leads to the simplification $A.B.\overline{C}.\overline{D} + A.B.C.\overline{D} = A.B.\overline{D}$. In the next figure 12.17(b), wrapping around both horizontal and vertical axes simplifies the map to $\overline{B}.\overline{D}$; all combinations of A and C appear and may be summed.

If all entries of the Karnaugh map are 1, what is the result?

Σ *notation*

An expression like equation (12.7) is called a **sum of products**. When written like this in terms of the basic elements of the truth table, it is said to be in **standard form**. Each term in the expression is called a **minterm**. There is a shorthand notation for an expression like equation (12.7). Suppose the binary number $ABCD$ is converted

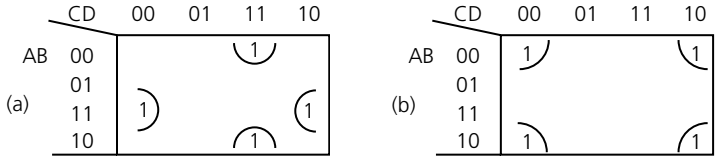


Fig. 12.17. Closing the Karnaugh map top and bottom and at the sides.

Copyright © 2005 IOP Publishing Ltd.

AB \ CD	00	01	11	10
00	0	1	3	2
01	4	5	7	6
11	12	13	15	14
10	8	9	11	10

Fig. 12.18. Decimal notation for entries of the Karnaugh map.

to decimal; the entries of the Karnaugh map may be numbered in decimal as shown in figure 12.18. Then the entries in figure 12.17(a), for example, may be designated by $\Sigma(3,11,12,14)$, where the Σ stands for a sum of products.

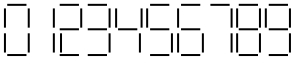
12.7 Don't Care or Can't Happen Conditions

Suppose a counter works in binary coded decimal (BCD). It counts from 0 to 9 in binary, then resets to 0. Suppose also that the numbers are to be displayed as shown in figure 12.19(a) on segments of light emitting diodes (LEDs). A Karnaugh map may be drawn for every segment of this display in terms of the four binary digits of the BCD number (exercise 9).

As an illustration, the logic required to light up the top left-hand vertical segment will be derived. This segment occurs in the numbers 0, 4, 5, 6, 8 and 9. The BCD number is represented by the four binary digits A, B, C and D and the required logic is $\Sigma(0, 4, 5, 6, 8, 9)$. However, a further simplification is possible. The numbers 10 to 15 cannot arise (unless the counter fails). In figure 12.19(b), these 'can't happen' conditions are marked by X's. We can choose to make them l's if it simplifies the Karnaugh map or we can choose to ignore them if it does not. Setting them to 1, the blocks ringed in the figure give the result $A + \overline{C}.\overline{D} + B.\overline{C} + B.\overline{D}$.

A second example arises in the logic of traffic lights, which should follow the sequence red → red + yellow → green → yellow → red. This example actually involves holding the current values of red(R), yellow(Y) and green(G) in a memory and will be discussed more fully in the next chapter. However, for the moment

(a)



AB \ CD	00	01	11	10
00	0			
01	4	5		6
11	X	X	X	X
10	8	9	X	X

Fig. 12.19. (a) Display of decimal numbers, (b) the Karnaugh map for the top left-hand vertical segment.

Copyright © 2005 IOP Publishing Ltd.

	RY	00	01	11	10	
	0	X	(1)	0	(1)	R_{new}
	1	0	(X)	X	(X)	
G	0	[X]	0	0	[1]	Y_{new}
	1	[1]	X	X	[X]	
	0	X	0	(1)	0	G_{new}
	1	0	X	(X)	X	

Fig. 12.20. Logic of traffic lights, allowing for 'can't happen' conditions, labelled X.

we need only consider the combinational logic giving new values of red, yellow and green when a clock pulse (C) triggers a change. These conditions are shown in figure 12.20. For example, if the current state is red, entries are at the top right corner of each rectangle ($R.\overline{Y}.\overline{G}$) and the new state is red + yellow: $R_{new} = 1$, $Y_{new} = 1$, $G_{new} = 0$. In the table, X indicates a condition which should not arise, e.g. red + yellow + green. In working out the logic, X may be counted as 1 or 0, whichever gives the more convenient result. This allows simplifications in the logic, which becomes

$$R_{new} = (\overline{R}.Y + R.\overline{Y}).C$$
$$Y_{new} = \overline{Y}.C$$
$$G_{new} = R.Y.C.$$

Of course, we might well decide that faulty combinations are not really 'don't care' but require special action, rather than being ignored. That would lead to different logic.

12.8 Products of Karnaugh Maps

There is one further trick which is useful in manipulating Karnaugh maps. Suppose we have a complicated product of expressions, such as

$$F = (\overline{A} + B + C).(\overline{B} + A.B + C.D).(\overline{A}.B + \overline{C}.D).(\overline{B} + A.C + D).$$

Multiplying this out is tedious and prone to errors. There is an alternative shown in figure 12.21.

When any two brackets are multiplied, the resulting Karnaugh map contains a 1 only where both brackets contain a 1. For example, the maps for the first two brackets are shown in figures 12.21(a) and (b). Each contains the term $\overline{A}.\overline{B}.\overline{C}.\overline{D}$ in the top left-hand corner. When we AND the two brackets, this term gives $\overline{A}.\overline{B}.\overline{C}.\overline{D}.\overline{A}.\overline{B}.\overline{C}.\overline{D} = \overline{A}.\overline{B}.\overline{C}.\overline{D}$. But the other two factors contain a zero there. Thus, the Karnaugh map of the expression F may be formed by drawing the maps of all 4 individual brackets then using the result that 1's are present in F only at points of the map where 1's are present in all four brackets. Another way of saying

Copyright © 2005 IOP Publishing Ltd.

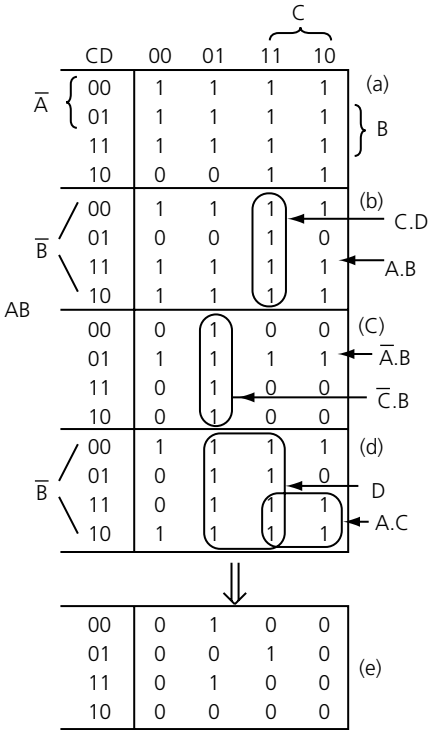


Fig. 12.21. Karnaugh maps for (a) $(\overline{A} + B + C)$, (b) $(\overline{B} + A.B + C.D)$, (c) $(\overline{A}.B + \overline{C}.D)$, (d) $(\overline{B} + A.C + D)$ and (e) their product.

this is that there is a 0 in the result whenever there is a 0 in any of the individual brackets. The result is illustrated in figure 12.21(e). Only three locations have a 1 common to all brackets, so the result is

$$F = \overline{C}.D.(\overline{A}.\overline{B} + A.B) + \overline{A}.B.C.D.$$

This may be verified by multiplying out F using Boolean algebra.

At this point, you are advised to get plenty of practice with Karnaugh maps by tackling exercises 1 to 9.

12.9 Products of Sums

It may sometimes be convenient to work out logical expressions in terms of the zeros of the Karnaugh map, rather than the ones. This can happen, for example, if there are rather few zeros, but a second practical reason will appear below. To illustrate the point, suppose figure 12.12 is redrawn plotting the zeros rather than the 1's, figure 12.22. The simplified expression for these locations is $\overline{A}.\overline{C} + \overline{A}.\overline{B} + \overline{B}.\overline{C}$.

Copyright © 2005 IOP Publishing Ltd.

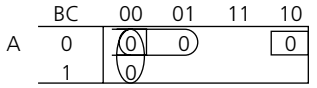


Fig. 12.22. Zeros for the carry signal of the full adder.

Since this expression stands for the *zeros* of the Karnaugh map, the expression for the carry signal can be rewritten as its complement:

$$\overline{\text{Carry}} = \overline{A}.\overline{C} + \overline{A}.\overline{B} + \overline{B}.\overline{C}. \tag{12.8}$$

This looks more complicated than before, until we realise that the expression may be simplified using de Morgan's theorem. Remember that this theorem states that every signal can be turned into its complement by interchanging AND with OR:

$$\text{Carry} = (A + C).(A + B).(B + C). \tag{12.9}$$

This form is known as a **product of sums** or alternatively as an OR–AND form. As a little exercise, this expression is multiplied out to verify that it gives the same result as equation (12.3):

$$\begin{aligned}\text{Carry} &= (A.A + A.C + A.B + B.C).(B + C)\\ &= (A + A.C + A.B + B.C).(B + C)\\ &= (A + B.C).(B + C) \qquad \text{using rule (vi) of Boolean algebra}\\ &= A.B + B.C + A.C\end{aligned}$$

which agrees with equation (12.3).

Equation (12.9) is a simplified expression. In standard form it reads

$$\begin{aligned}\overline{\text{Carry}} &= \overline{A}.\overline{B}.\overline{C} + \overline{A}.B.\overline{C} + A.\overline{B}.\overline{C} + \overline{A}.\overline{B}.C\\ \text{Carry} &= (A + B + C).(A + \overline{B} + C).(\overline{A} + B + C).(A + B + \overline{C}).\end{aligned}$$

The individual terms in this expression, such as (A + B + C), are known as **max-terms** or **standard sums**. An alternative notation for the result is:

$$\text{Carry} = \Pi(0, 1, 2, 4);$$

this stands directly for the locations of the zeros in the Karnaugh map, and the symbol Π indicates that the product is to be formed of the corresponding sums to which they are related by de Morgan's theorem.

12.10 Use of NOR and NAND Gates

Some families of logic gates are only readily available in the form of NAND and NOR gates. This raises the question of how to go about using inverted gates. This

Copyright © 2005 IOP Publishing Ltd.

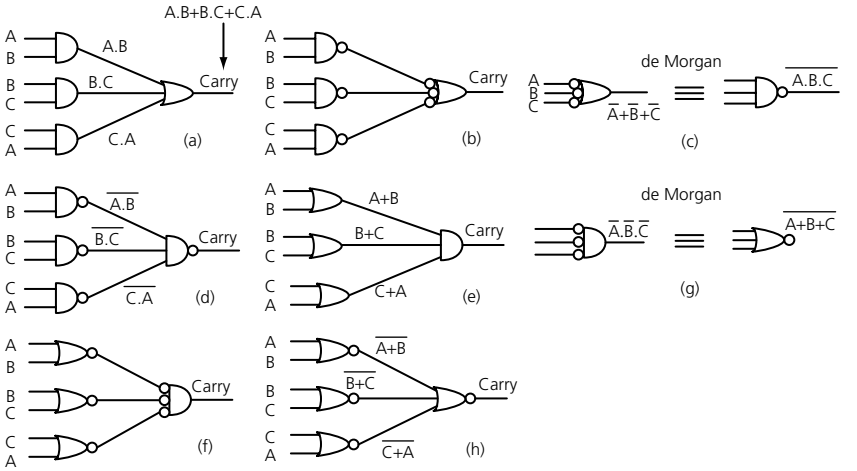


Fig. 12.23. (a), (b), (d), (e), (f) and (h): various realisations of the carry of the full adder; (c) and (g) diagrammatic representations of de Morgan's theorems.

is actually easier than it may appear. Figure 12.23(a) shows the logic for forming the Carry signal of a full adder from equation (12.3) in AND-OR form. Suppose every signal is inverted twice between the gates; this is indicated by the bubbles in figure 12.23(b). It has no overall effect. Now de Morgan's theorem states that inputs and outputs may be complemented if OR is changed to AND; as a reminder, this is shown diagrammatically in figure 12.23(c) and is usually the simplest way of using this theorem. Figure 12.23(b) can then be redrawn as in (d). This circuit achieves the same result as figure 12.23(a), but using NAND gates only. It is called a NAND-NAND form. The diagrammatic procedure shown here is quick and simple.

Let us now examine the relevance of the product of sums representation. Equation (12.9) leads to the circuit of figure 12.23(e). Figures 12.23(f) to (h) then show how to convert this into a form using only NOR gates. This is called a NOR-NOR form. As an algebraic check, figure 12.23(h) forms $\overline{\overline{A+B}+\overline{B+C}+\overline{C+A}} = (A+B).(B+C).(C+A)$ by de Morgan's theorem.

The conclusion is that sums of products (derived from the 1's of the Karnaugh map) are convenient for obtaining a circuit in terms of NAND gates, while products of sums (derived from the 0's) are the starting point for obtaining logic in terms of NOR gates. However, a word of warning is appropriate here. Positive logic is much easier to follow and test. It may be worth a small extra expense or an extra gate or two to be able to express logic in terms of AND and OR gates when making a single circuit or a few off. The economies achieved with NAND and NOR gates may be essential on a large scale, and it is necessary to be conversant with the manipulations illustrated in figure 12.23 and to be able to decipher the logic diagrams.

Copyright © 2005 IOP Publishing Ltd.

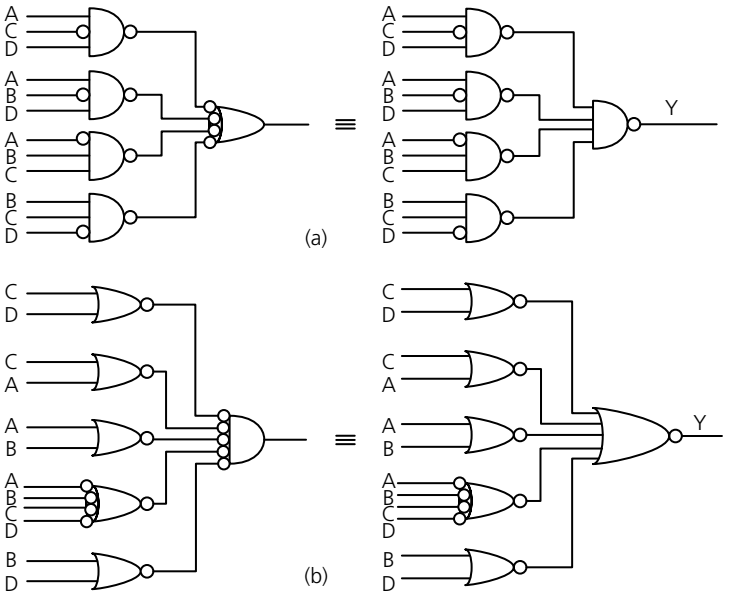


Fig. 12.24. Realisation of Y of figure 12.16(b) in terms of (a) NAND and (b) NOR gates.

As a second example, Y of Figure 12.16(b) is given by $A.\overline{C}.D + A.\overline{B}.D + \overline{A}.B.C + B.C.\overline{D}$ and hence by the NAND gate configuration of figure 12.24(a). Alternatively, the zeros of figure 12.16(b) are given by

$$\overline{Y} = \overline{C}.\overline{D} + \overline{C}.\overline{A} + \overline{A}.\overline{B} + A.B.C.D + \overline{B}.\overline{D}$$

or $$Y = (C + D).(C + A).(A + B).(\overline{A} + \overline{B} + \overline{C} + \overline{D}).(D + B)$$

and the NOR gates of figure 12.24(b).

This is a good point to get some practice in manipulating gates, exercises 10 to 16.

12.11 Decoders and Encoders

So far logic has been assembled out of simple gates. There is a standard unit which is used in large-scale logic as a convenience. This is the decoder. Two examples are shown in figure 12.25. When enabled by a signal on line E, the device in (a) gives an output on one of the four lines $D0$ to $D3$ according to the value of the binary input AB. For example, if A and B are both 1, an output 1 appears on line $D3$. Likewise, (b) decodes a 3-digit binary number to give an output on one of the 8 output lines. The two units work by simply forming all possible minterms, one for each output line, as in figure 12.25(c).

Copyright © 2005 IOP Publishing Ltd.

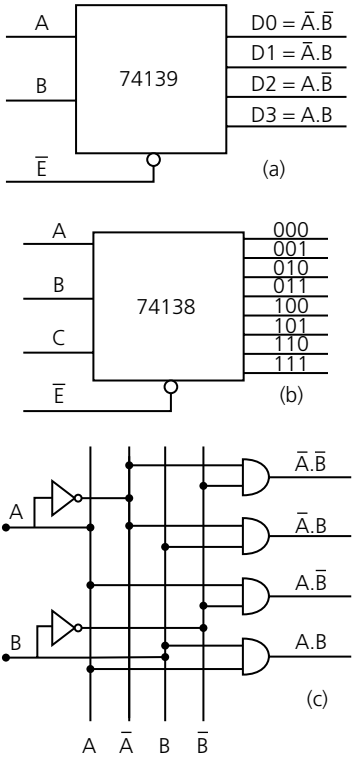


Fig. 12.25. (a) 2 to 4 line decoder, (b) 3 to 8 line decoder, and (c) the logic of (a).

Such devices are a very convenient way of doing sum of products logic since all minterms of the Karnaugh map are available as outputs. Suppose, for example, a full adder is to be constructed. For this, the Sum signal can be written, figure 12.14, Sum = $\Sigma(1, 2, 4, 7)$ and the Carry, figure 12.12, as $\Sigma(3, 5, 6, 7)$. These can be formed by feeding the 74138 decoder with inputs A, B and C and taking the required combination of the output lines to two OR gates, one for Sum (1 + 2 + 4 + 7) and one for Carry (3 + 5 + 6 + 7). When line E is enabled, the outputs of the two OR gates generate Sum and Carry. Output line $D0$ of the decoder is unused, but could for example identify zero input. Another obvious use for the decoder is to address registers or devices in a computer.

Encoders are devices working the other way round. For example, a decimal to binary converter has 10 input lines plus an enable, and 4 output lines corresponding to the four binary digits. A keyboard encoder has 84 or so input lines, corresponding to the number of symbols (upper and lower case combined) on the keyboard and 7 output lines for binary output code.

A technical detail is that encoders and decoders are usually activated by the enable signal going LOW rather than HIGH. In circuit diagrams, this is denoted

Copyright © 2005 IOP Publishing Ltd.

by a bubble on the enable line, to indicate a complementary signal. Such an input control line is said to be **active low**. The circuit is permanently enabled if the line E is earthed.

12.12 Multiplexing

Suppose we have two slow streams of data, A and B. They can be gated alternately on to a single faster line using a control pulse C by forming $A.C + B.\overline{C}$: A goes on to the line when $C = 1$ and B goes on when $C = 0$. Then they can be unscrambled at the other end of the line by using the C pulse to route A and B to separate outputs:

$$A = C.(A.C + B.\overline{C})$$

$$B = \overline{C}.(A.C + B.\overline{C}).$$

More generally, many lines can be multiplexed on to a single line by making C an address from which the input is selected and to which the output is sent. As an example, a number of telephone conversations may be routed on to a single fast line by sampling each conversation at a rate above audio frequencies and then unscrambling them at the other end of the line. This technique is called **multiplexing** and the device which does it is called a **multiplexer**. A 4 to 1 line MUX is shown in figures 12.26(a) and the way it works in (b). It is essentially an encoder with a single output line. The decoding system is called a **demultiplexer** (DEMUX); this is achieved simply by connecting the fast line to the enable input of a decoder and the appropriate address to the select lines (A and B of Figure 12.25(a)).

Multiplexing is used extensively in data processing systems where many registers (or devices) may feed on to a common bus line; encoders and decoders arrange that each register may be addressed individually.

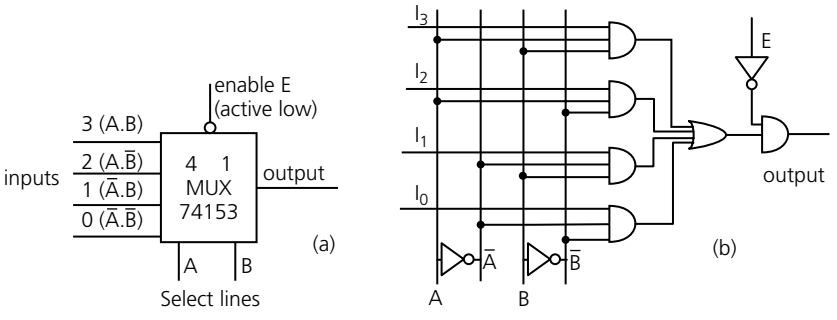


Fig. 12.26. (a) Symbolic for m and (b) gate layout of a 4 to 1 multiplexer.

Copyright © 2005 IOP Publishing Ltd.

12.13 Exercises

1. Convert decimal 23 and 39 to binary and add them together in this systems. Write the representation of 56.1_{10} in binary and BCD. (Ans: $10111 + 100111 = 111110$; 111000.000110011 …, 01010110.0001.)

2. Give truth tables for $X = \overline{A} + B$, $Y = \overline{A}.B.C$, and $Z = A \oplus B + \bar{C}$. Write the truth table for NOR of A and B and show that it is the same as $\overline{A}.\overline{B}$.

3. (QMC). Explain, with the aid of truth tables, the meaning of the logical operations AND, OR, NOT, NOR, NAND. Give the logic circuit symbols for each and sketch the two-input transistor-resistor circuits that would implement these operations. Verify using truth tables that a NAND gate with the inputs connected together is a NOT gate and that a NAND gate with the inputs inverted is an OR gate.

4. (QMC). The circuit of Figure 12.27(a) can be used to perform a logical operation on digital pulses applied to inputs A and B. A signal of $+V$ volts is used for the logical '0' state and zero volts for the logical '1' state. The resistors R_2 have values several times larger than R_1. Give the name of the logical operation and its truth table. Describe how the circuit works. How would you represent the logical action of this circuit using switches? Draw the output that would result from input pulses shown in (b). If now the pulses are redefined so that $+V$ becomes the 'l' state and zero volts the '0' state, what is the new function of the circuit? Draw a truth table for the inputs and output. (Ans: OR, Figure 12.27(c), AND.)

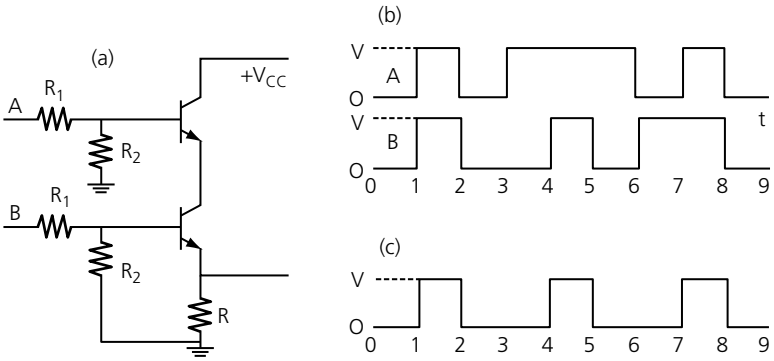


Fig. 12.27.

5**. (QMC). For Figure 12.28, derive a simple form for the output f. What is this logic function? With an inverter used to derive $\overline{y}$ from an input y, carry out a waveform analysis of the circuit. (Ans: $x.\overline{y} + \overline{x}.y$, XOR.)

Copyright © 2005 IOP Publishing Ltd.

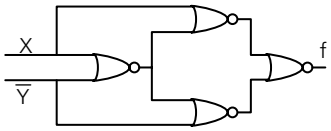


Fig. 12.28.

6. (RHBNC). Take the Boolean expression $f = A.\overline{B}.\overline{C} + A.\overline{B}.C$ and simplify it using a Karnaugh map. Also simplify it algebraically and hence show that both methods are equivalent. (Ans: $f = A.\overline{B}$.)

7. (QEC). (a) Simplify the following logical expressions: (i) $A.B + A.\overline{C} + \overline{A}.\overline{B}.\overline{C}$, (ii) $(\overline{A}+\overline{B}+C).(A.B+D)$, $*$(iii) $A+A.B+\overline{A}.C.\overline{D}+\overline{A}.\overline{B}.C.D+B.C.D$; (b) Use a Karnaugh map to minimise the following expressions: (i) $\overline{A}.\overline{B}.\overline{C}+\overline{A}.B.C+A.B+A.\overline{B}.\overline{C}+A.C$; (ii) $C.\overline{D}+A.C.D+\overline{B}.\overline{C}.\overline{D}+A.\overline{B}.D + \overline{A}.B.\overline{C}.\overline{D}$. (Ans: (a) $A.B + \overline{B}.\overline{C}$, $A.B.C + (\overline{A} + \overline{B} + C).D$, $A + C$; (b) $A + B.C + \overline{B}.\overline{C}$ and $A.\overline{B} + \overline{A}.\overline{D} + A.C$.)

8. (QMC). Use the Karnaugh map to obtain the simplest minimal forms of the following functions: $f_1 = A.\overline{B}.\overline{C}.D + A.\overline{B}.C + A.\overline{B}.\overline{C}.\overline{D} + \overline{A}.\overline{B}.C.D + A.\overline{B}.C.\overline{D}$, $f_2 = \overline{A}.\overline{C}.\overline{D} + \overline{A}.B.\overline{C} + \overline{B}.\overline{C}.\overline{D} + \overline{A}.B.C + A.B.C$, $f_3 = (A+B+\overline{D}).(B+C+\overline{D}).(\overline{A}+B+\overline{C}+\overline{D}).(A+\overline{B}+\overline{D})$. What is the practical importance of this minimisation? (Ans: $f_1 = \overline{B}.(A + C.D)$, $f_2 = \overline{A}.B + \overline{B}.\overline{C}.\overline{D} + B.C$, $f_3 = A.B + \overline{D}$.)

9. (RHBNC). The seven segment LED display showing a BCD count is shown in Figure 12.19(a). Design a combinational logic circuit to drive its three horizontal bars. Take the binary number to be ABCD, as in figure 12.18. (Ans: top $= A + D.(B + C) + \overline{B}.\overline{D}$, middle $= A + C.(\overline{B} + \overline{D}) + B.\overline{C}$, bottom $= C.(\overline{B} + \overline{D}) + \overline{B}.\overline{D} + B.\overline{C}.D$).)

10*. (RHBNC). You are provided with a square wave generator G and a divide-by-eight counter which triggers on the trailing (negative going) edge of G; the counter has outputs A, B and C for 1, 2 and 4. Using these and their complements and NAND logic gates only, design a circuit which will provide the output waveform in figure 12.29, assuming that the counter is initially reset to zero before the first G pulse arrives. (Positive logic when the output and G are high). (Ans: Figure 12.30.)

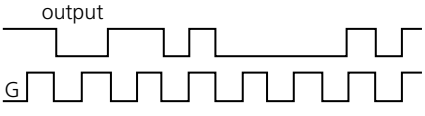


Fig. 12.29.

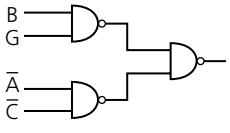


Fig. 12.30.

Copyright © 2005 IOP Publishing Ltd.

11*. (RHBNC). Design the circuit of a comparator for two 1-bit numbers. It should output 1 if the two bits are identical and zero otherwise. (Ans: XNOR $= A.C + \overline{A}.\overline{C}$.)

12*. Show both algebraically and by rearranging OR and NAND gates that the exclusive OR $(A \oplus B)$ can be expressed in the following 8 forms: $A'.B + A.B'$, $(A.B + A'.B')'$, $(A + B).(A' + B')$, $((A'B)'.(A.B')')'$, $((A + B)' + (A' + B')')'$, $(A.B)'.(A'.B')'$, $(A + B')' + (A' + B)'$ and $((A + B').(A' + B))'$, where $'$ denotes the complement.

13. Show that $\overline{A}.B + C.D = \Sigma(3,4,5,6,7,11,15)$. Express it also as a product of maxterms. How could you form this result using two 2-bit decoders? (Ans: $\Pi(0,1,2,8,9,10,12,13,14)$.)

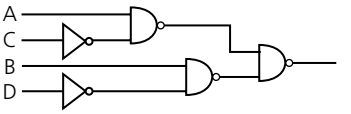


Fig. 12.31.

14. (RHBNC). Draw a Karnaugh map for the sum of products function $f = \Sigma(4,6,8,9)$ in the range 0 to 11. Minimise this and draw the final circuit using NAND and NOT gates. Take inputs to be A, B, C, D. (Ans: $B.\overline{D} + A.\overline{C}$; Figure 12.31.)

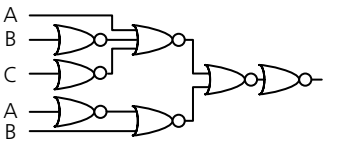


Fig. 12.32.

15**. (Bedford). Give the truth tables for the logical expressions AND, OR, NOT and NOR. State de Morgan's theorems and show that the NOR operation can be used as a universal decision function to replace the three primary operations specified above. Draw a logic circuit diagram to show how the function $f = (A + B.C).(B + \overline{C}).\overline{A} + A.\overline{B}$ may be implemented using NOR gates, assuming that the variables A, B and C are available as inputs. (Ans: Figure 12.32.)

Copyright © 2005 IOP Publishing Ltd.

13

Sequential Logic

13.1 The RS Flip-Flop

A flip-flop is a combination of gates having two alternative stable states which can represent 0 and 1. A trigger pulse can flip it from one state to the other. Until this trigger pulse arrives, the circuit maintains its logical state indefinitely. The logic gates discussed in the previous chapter were capable of making a decision; the flip-flop memorises the decision.

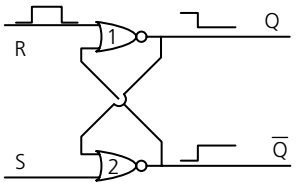


Fig.13.1. The RS flip-flop and waveforms when a reset is applied.

S	R	Q	$\bar{Q}$
0	0	1	0
0	0	0	1
1	0	1	0
0	1	0	1

(a)

S	R	Q(t + 1)
0	0	0
0	1	0
1	0	1
1	1	?

(b)

Fig.13.2. (a) State table, (b) excitation table of the RS flip-flop.

The simplest variety is shown in figure 13.1. The convention will be adopted here that a logical 0 is represented by a low level and logical 1 by a high level (positive logic). Consider first the voltage levels in the absence of any input pulses and suppose that gate 1 gives an output Q which is positive (high). This signal drives gate 2 so that its output $\overline{Q}$ is low (logical 0). This gate feeds a signal back to gate 1 which is low. In the absence of an input from A, the output Q is high. The inversion of the signal at both gates provides overall **positive feedback**: the signal fed back to the input of each gate reinforces the existing output. Because of the symmetry of the circuit, there is a second stable configuration with the opposite polarity: gate 1 low and gate 2 high.

As long as there is no external input to either gate, either configuration is stable. However, if a large enough **reset** pulse R is applied to gate 1, it drives Q low and $\overline{Q}$ high. The corresponding waveforms are sketched in the figure. A positive **set** pulse S at gate 2 restores the original situation, i.e. Q high, $\overline{Q}$ low.

Copyright © 2005 IOP Publishing Ltd.

State and Excitation Tables

This sequence of events can be traced through the two tables of figure 13.2. The first shows the quiescent states with no input to S or R, or at most a 1 on one of them. This is called the **state table** of the flip-flop. The second is the **excitation table** showing the changes between the output Q at some time and the output $Q(t+1)$ after switching. With a 1 on input R (reset) and a 0 on input S, the flip-flop goes to $Q(t+1) = 0$ whatever its previous condition; conversely, with a 1 on input S (set) and a 0 on input R, it goes to $Q(t+1) = 1$. In the absence of any input, (i.e. $S = R = 0$), it stays steady. The device is called a **set-reset (RS) flip-flop** or **latch**. If it is fed contradictory information, a 1 on both R and S input lines, both Q and $\overline{Q}$ temporarily go to 0; however, this is not a stable configuration, so the circuit settles eventually according to whichever input survives longest.

Debounce circuit

An elementary example using this flip-flop is the **debounce** circuit shown in figure 13.3. Suppose a piece of electronics is to change state under the action of a mechanical switch. When this switch is moved from position S to position R, the contacts may make and break several times at R before settling to a good contact. It is desirable that the electronics should respond to the first contact and then remain stable, rather than switching back and forth as the circuit makes and breaks. This is achieved by the flip-flop of figure 13.3(a), which is reset to $Q = 0$ by the first 1 signal on line R and remains in a fixed state until the switch is moved back to position S, when a 1 signal appears on line S and sets the flip-flop to $Q = 1$.

13.2 Clocks

In a large and complicated system like a computer, it is easy for small timing differences to appear between different parts of the circuitry. Signals propagate just below the speed of light, 30 cm/ns. Switching times of 1 or 2 ns are realistic in high-speed computers, so timing is a serious problem.

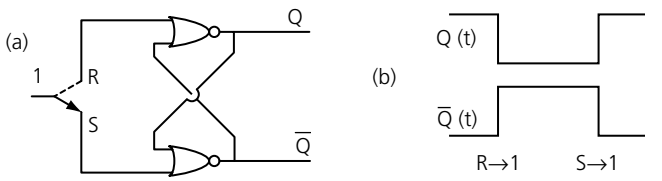


Fig. 13.3. (a) A debounce circuit and (b) its waveforms.

Copyright © 2005 IOP Publishing Ltd.

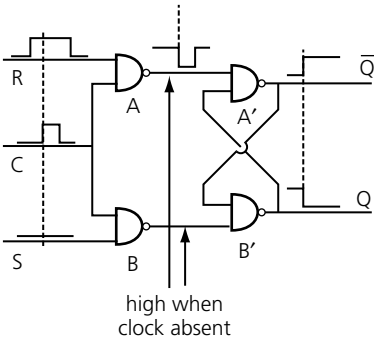


Fig. 13.4. Clocked RS flip-flop.

For this reason, it is common to synchronise the switching of gates by a **clock** or **strobe** pulse. To achieve this, it is desirable to reorganise the flip-flop as in figure 13.4 so that it uses NAND gates rather than NOR; then the clock pulse C can be a pre-requisite for the NAND gates A and B to respond. In the absence of a clock pulse, both gates A and B give high outputs, so both NAND gates A' and B' of the flip-flop itself are enabled and can respond to the feedback between Q and $\overline{Q}$. You may wish to work through the state table and excitation table, to follow the quiescent logic and the switching. The waveforms on the figure show switching from $Q(t) = 1$ to $Q(t + 1) = 0$. The clock pulse C is made narrower than the R pulse and synchronises the switching, as indicated by the dashed lines. Logic like this, regulated by a clock pulse, is called **synchronous**. Logic operating without clock pulses is called **asynchronous**.

The RS flip-flop suffers the disadvantage that its output is not well defined if $S = R = 1$. A way of avoiding this contradictory situation is shown in figure 13.5. The S and R inputs are derived from a single input D using an inverter to create R from S. This is known as a **D flip-flop**, where D stands for data: the state of the flip-flop is dictated by the data fed to its input. If $D = 1$ it sets to 1 when activated by a clock pulse; if $D = 0$ it resets to 0.

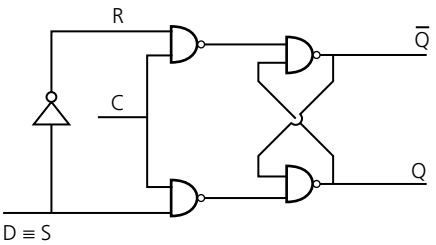
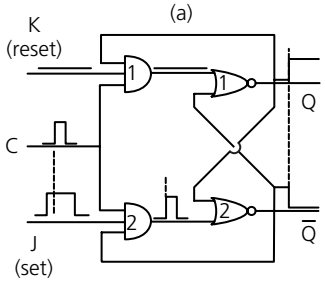


Fig. 13.5. The D flip-flop.

Copyright © 2005 IOP Publishing Ltd.



(b)

Q(t)	J	K	Q(t + 1)
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	1
1	0	0	1
1	0	1	0
1	1	0	1
1	1	1	0

Fig. 13.6. (a) A primitive JK flip-flop and (b) its excitation table.

13.3 The JK Flip-Flop

The RS flip-flop does not respond in a well defined way to $R = S = 1$. A step towards solving this problem is given in figure 13.6(a). Here, the Q output is fed back at the top of the figure to the input AND gate 1, so as to enable this gate when $Q = 1$; likewise, the $\overline{Q}$ output is fed back to the other input AND gate 2, enabling it only when $Q = 0$. Because Q and $\overline{Q}$ are opposite logic states, only one of the AND gates can open and the switching action is well defined, as given in figure 13.6(b). If $Q = 0$, only the J input can do anything; if $Q = 1$, only the K input has any effect.

The upshot is that J alone acts as a set input, K alone acts as a reset, and if both inputs are present, the flip-flop changes state, or 'toggles', via whichever AND gate is currently open.

Unfortunately, there is a snag with figure 13.6(a), and it cannot be used as it stands. Suppose the clock pulse is long compared with the switching time of the flip-flop. In this case, as soon as the Q and $\overline{Q}$ outputs change state, the enabling of the input AND gates swops from J to K or vice versa. If $J = K = 1$ and the clock pulse is still there, the flip-flop will toggle back again to its original state. It may continue to do this several times until one of C, J or K inputs disappears; the final state is ill-defined.

Master Slave Flip-Flop

An alternative solution is shown in figures 13.7(a) and (b). In this arrangement, there are two flip-flops, called master and slave. The first master flip-flop responds to the clock pulse and may change state if J and K are set appropriately; the second slave flip-flop accepts the output of the master flip-flop only when the clock pulse disappears. Consequently, the feedback of Q and $\overline{Q}$ to the input AND gates can never arrive until the clock pulse has disappeared, and multiple toggling is locked out. The waveforms of figure 13.7(c) describe the response to an input $K = 1$ (reset) starting from $Q = 1$. In this case, the switching action of the slave is initiated at gate 5 by $\overline{C}$ going high with $Y = 1$.

Copyright © 2005 IOP Publishing Ltd.

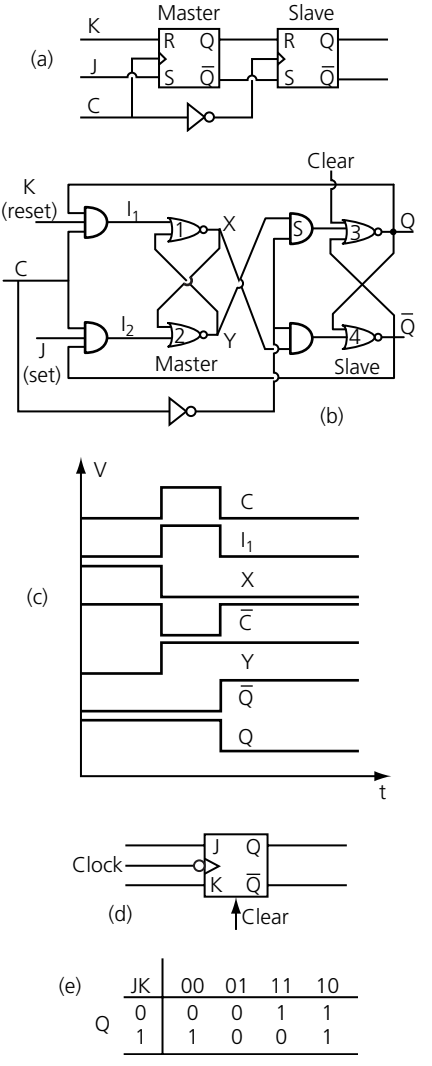


Fig. 13.7. The master-slave JK flip-flop: (a) schematic, (b) gate layout, (c) waveforms when J = K = 1 and Q is initially 1, (d) symbolic form, (e) Karnaugh map for Q(t+1).

The switching action of the flip-flop is designed to respond to the input pulse changing from low to high or vice versa. Response to an input changing from low to high is called **positive edge triggering**. Response to the reverse transition from high to low is known as **negative edge triggering**. In figure 13.7(d), the bubble on the clock input to the flip-flop indicates negative edge triggering.

Commercial chips usually have the additional feature, shown in (d), of an extra input which will clear the flip-flop to $Q = 0$ even without a clock pulse. This allows a register of flip-flops to be cleared initially to zero. The clear input bypasses the

Copyright © 2005 IOP Publishing Ltd.

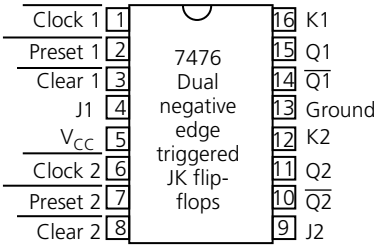


Fig. 13.8. Layout of IC 7476.

initial AND gates of figure 13.7(b) and is fed directly to NOR 3. The layout of the 7476 JK flip-flop is shown in figure 13.8, and is convenient if you wish to play in the laboratory with circuits shown in this chapter.

A feature of the *JK* flip-flop is that with $S = R = 1$ it toggles from $Q = 1$ to 0 or vice-versa. This makes it ideal for use in counters. A version is available with the J and K lines permanently connected together. If $J = K = 0$ it does nothing; if $J = K = 1$ it changes state. Such a device is called a **Toggle (T) flip-flop**.

The equation for the JK flip-flop

In following the action of the *JK* flip-flop, it will be important later to use its **characteristic equation**:

$$\boxed{Q(t+1) = J.\overline{Q} + \overline{K}.Q.} \tag{13.1}$$

This expresses the state $Q(t+1)$ after switching in terms of the initial state Q and the inputs J and K. It may be derived from the Karnaugh map of figure 13.7(e) and may be checked against the excitation table of figure 13.6(b). This equation may be used systematically to devise switching systems.

13.4 A Scale-of-4 Counter

In order to appreciate how to use flip-flops, let us consider as an example how to make a counter which counts from 0 to 3 in binary and then resets. Suppose A represents its more significant digit and B the less significant one. The switching action required in response to clock pulses is shown in figures 13.9(a) and (b). Using set and reset signals applied to the A and B flip-flops, the required equations are:

$$S_B = \overline{B} \qquad R_B = \overline{S}_B$$

$$\text{and} \qquad S_A = A \oplus B \qquad R_A = \overline{S}_A.$$

Copyright © 2005 IOP Publishing Ltd.

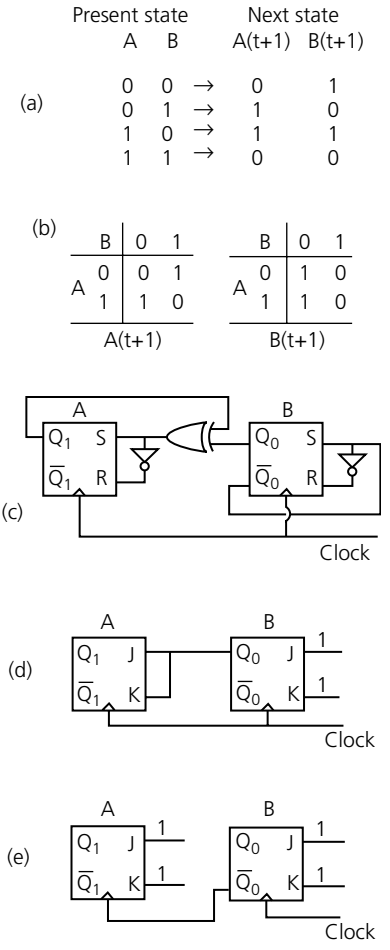


Fig. 13.9. Design of a scale-of-4 counter: (a) excitation table, (b) Karnaugh maps, (c) circuit using RS or D flip-flops, (d) circuit using JK or T flip-flops, (e) ripple-through carry.

A circuit to achieve this with D flip-flops is shown in figure 13.9(c). Alternatively, because the JK flip-flop toggles when $J = K = 1$, the circuit of figure 13.9(d) achieves the same result without the XOR gate required in figure 13.9(c).

If clock pulses are fed in at a constant rate, the Q output of flip-flop B goes positive at half the rate of input clock pulses; $Q(A)$ goes positive at a quarter of the rate. Counters can therefore be used for frequency division.

In the circuits of figures 13.9(c) and (d), both digits A and B change in synchronism with the clock pulse. A less satisfactory alternative is shown in figure 13.9(e), where the clock pulse for A is provided by the positive edge given by $\overline{Q}_0$ switching to 1. In this case, A switches only after the propagation delay through the less significant digit B. If the counter has many digits, and is in the state ... 1111,

Copyright © 2005 IOP Publishing Ltd.

the carry propagates through each of the lower digits in turn, with a delay in each. This is called **ripple-through carry**. This mode of operation may be quite satisfactory at low speeds, but is undesirable in a high-speed counter, since input pulses may be arriving at a rate comparable with the rate of carry propagation, and at any instant the counter will not be giving an up-to-date reading; it would give a false result if it drives further logic. If the input pulse train is halted, the counter will catch up and give a correct reading after the carries have propagated through.

Bidirectional counter

Next, suppose the counter is to count forwards if a control signal X is 1 but backwards if $X = 0$. This makes a **bidirectional counter**. The excitation table and Karnaugh maps are given in figures 13.10(a) and (b). From (b), $B(t+1) = \overline{B}$. Comparing this with the characteristic equation (13.1) of the JK flip-flop, which reads $B(t+1) = J.\overline{B} + \overline{K}.B$, we need $J = 1$ and $\overline{K} = 0$, i.e. $K = 1$. This result is intuitively obvious from figure 13.10(a), which requires B to toggle every time. Next (b) gives

$$A(t+1) = A.(B\overline{X} + X.\overline{B}) + \overline{A}.(\overline{BX} + B.X)$$

$$= A.(B \oplus X) + \overline{A}.(\overline{B \oplus X}).$$

Comparing with the JK characteristic equation $A(t = 1) = J.\overline{A} + \overline{K}.A$, this requires $J_A = \overline{B \oplus X}$, $\overline{K} = B \oplus X$, hence $K = \overline{B \oplus X} = J_A$. This is achieved with the circuit of figure 13.10(c).

(a)

Present state		Next state: X = 0 (backwards)		Next state: X = 1 (forwards)	
A	B	A	B	A	B
0	0	1	1	0	1
0	1	0	0	1	0
1	0	0	1	1	1
1	1	1	0	0	0

(b)

X	AB	00	01	11	10	
	0	1	0	1	0	A(t+1)
	1	0	1	0	1	

X	AB	00	01	11	10	
	0	1	0	0	1	B(t+1)
	1	1	0	0	1	

(c)

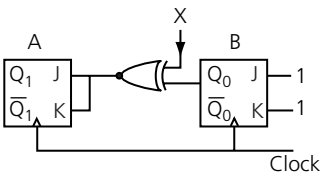


Fig. 13.10. Design of a bidirectional scale-of-4 counter: (a) excitation table, (b) Karnaugh maps, and (c) circuit using JK flip-flops.

Copyright © 2005 IOP Publishing Ltd.

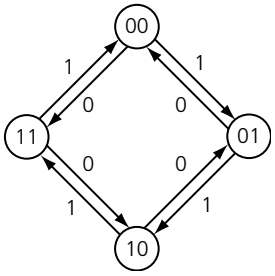


Fig. 13.11. State diagram for the bidirectional counter.

13.5 State Diagrams

A useful way of representing the switching action of the bidirectional counter is shown on figure 13.11. The steady states are shown in the circles, and the arrows indicate the direction of the changes when $X = 0$ or 1. This is called a **state diagram**.

As a second example, consider the switching of traffic lights, discussed in the previous chapter. There are four transitions:

$$\text{red} \rightarrow \text{red} + \text{yellow} \rightarrow \text{green} \rightarrow \text{yellow} \rightarrow \text{red}.$$

Again there are four quiescent states, which can be labelled a, b, c and d; the state diagram, figure 13.12, is equivalent to the scale-of-4 counter if these four states are represented either by

$$a = 00 \qquad b = 11 \qquad c = 10 \qquad d = 01$$

$$\text{or} \qquad a = 00 \qquad b = 01 \qquad c = 10 \qquad d = 11.$$

In either case, additional logic has to be provided to convert the output of the A and B flip-flops into a code driving the 3 colours of the traffic lights, e.g. $b \rightarrow$red + yellow.

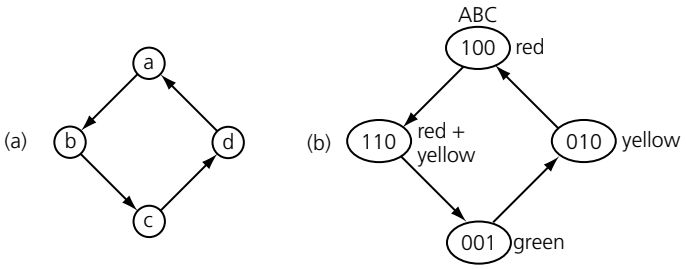


Fig. 13.12. (a) State diagram for traffic lights and (b) a possible digital representation.

Copyright © 2005 IOP Publishing Ltd.

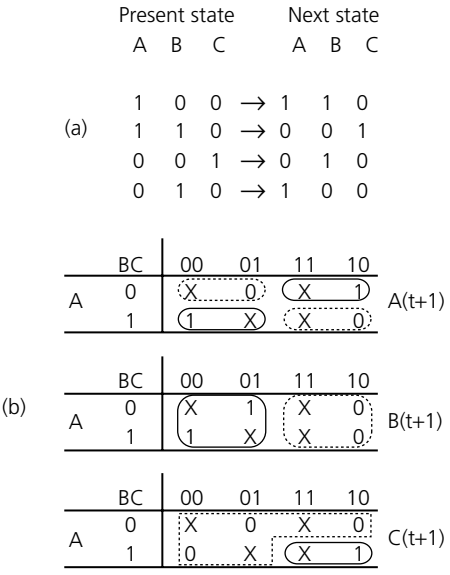


Fig. 13.13. (a) Excitation table and (b) Karnaugh maps for figure 13.12(b).

An alternative arrangement is instead to code the colours of the traffic lights directly in states a, b, c and d. For example, in figure 13.12(b), three digits ABC are used to code red, yellow and green directly: red = 100, yellow = 010, etc. The required excitation table and Karnaugh maps are shown in figures 13.13(a) and (b). The latter shows tables for $A(t+1)$, $B(t+1)$ and $C(t+1)$. These tables are constructed is as follows. Suppose the initial state is 100. From figure 13.13(a), the next state requires $A(t+1) = 1$, $B(t+1) = 1$, $C(t+1) = 0$. These values are entered into figure 13.13(b). Further initial states are treated likewise. Finally, X represents a don't care condition.

There are many possible ways to do the logic, but a simple choice using SR flip-flops would be:

$$S_A = A.\overline{B} + \overline{A}.B \qquad R_A = \overline{S}_A$$

$$S_B = \overline{B} \qquad R_B = B = \overline{S}_B$$

$$S_C = A.B \qquad R_C = \overline{S}_C.$$

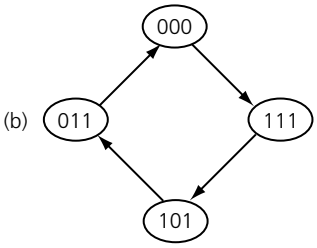
Four of the eight possible logical combinations of A, B and C do not occur in the state diagram of figure 13.12(b). What happens if the logic gets into one of these states? A power surge, for example, might temporarily set the combination yellow + green = 011. It is necessary to check that the system does not get stuck in the 4 unwanted combinations. Figure 13.14(a) shows that the logic chosen here leads in every case back to one of the required combinations. A system which returns to the required sequence whatever the initial state is said to be **self-starting**.

Copyright © 2005 IOP Publishing Ltd.

(a)

Present state				Next state		
A	B	C		A	B	C
0	0	0	→	0	1	0
0	1	1	→	1	0	0
1	0	1	→	1	1	0
1	1	1	→	0	0	1

(b)



(c)

Present state				Next state		
A	B	C		A	B	C
1	0	0	→	1	1	0
1	1	0	→	0	0	1
0	0	1	→	0	1	0
0	1	0	→	1	0	0
0	0	0	→	1	1	1
0	1	1	→	0	0	0
1	0	1	→	0	1	1
1	1	1	→	1	0	1

Fig. 13.14. (a) Excitation table for faulty states, (b) loop of faulty states, (c) the excitation table.

However, to illustrate a point, it is possible to derive characteristic equations which would drive the lights round the unwanted loop of 4 'illogical' states shown in figure 13.14(b), without ever returning to the desired states. Figure 13.14(c) gives the full excitation table for figures 13.13(b) and 13.14(b). The corresponding logic is

$$S_A = \overline{B}.\overline{C} + A.B.C + \overline{A}.B.\overline{C} \qquad R_A = \overline{S}_A$$

$$S_B = \overline{B} \qquad R_B = \overline{S}_B$$

$$S_C = A.B + \overline{A}.\overline{B}.\overline{C} + A.C \qquad R_C = \overline{S}_C.$$

It clearly pays to think about the 'don't care' states, and in this example it might be sensible to arrange that all unwanted states lead to yellow at the next change. This leads to a unique Karnaugh map with no don't care states.

Copyright © 2005 IOP Publishing Ltd.

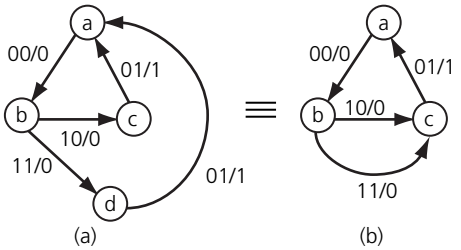


Fig. 13.15. Two equivalent trapping systems.

13.6 Trapping Sequences: Pattern Recognition

Another example is rather artificial in the simplified form given here, but is typical of a class of problems and illustrates several points. Consider a digital system which receives numbers at regular intervals governed by a clock. For simplicity, suppose these numbers are restricted to 0, 1, 2 or 3. A situation might arise where we want to recognise the sequences 0, 2, 1 or 0, 3, 1 and take some action. This is a **pattern recognition** problem. It might be used, for example, to trap faulty operation in a machine or a process.

The two sequences are represented on the state diagram of figure 13.15. Three states a, b, c can be used to represent the sequence 0, 2, 1. In figure 13.15(a), the first pair of digits on each arrow indicates the input pair of binary input digits which take the system through the required patterns $a \rightarrow b \rightarrow c \rightarrow a$. Likewise $a \rightarrow b \rightarrow d \rightarrow a$ recognises the second sequence of inputs 0, 3, 1. Any other pair of input digits restores the system to state a. The number after the stroke is 1 if the system is to produce an output.

The response of states c and d to any pair of input digits is identical. It is a general rule that in this case there is redundancy in the system, and one of the two states can be eliminated. This is done in (b) by eliminating state d and providing two alternative routes from b to c.

It is arbitrary how states a, b and c are coded. Suppose they are represented by flip-flops AB taking values $a = 00$, $b = 01$, $c = 11$. The excitation diagram and Karnaugh maps are shown in figure 13.16, where x and y represent the two input digits. Figure 13.16(a) is easily constructed by putting in entries corresponding to the four arrows of figure 13.15(b) and filling the remaining positions with zeros. There is an unused combination $AB = 10$, which should not occur. To deal with it, corresponding entries in the table may be set to 0, so that the system returns to state a.

The switching logic for A and B and the output C is:

$$A(t+1) = \overline{A}.B.x \tag{13.2}$$

$$B(t+1) = \overline{A}.B.x + \overline{A}.\overline{B}.\overline{x}.\overline{y} \tag{13.3}$$

$$C = A.B.\overline{x}.y.$$

Copyright © 2005 IOP Publishing Ltd.

(a)

Present state AB	x	y	Next state AB	Output C
a = 00	0	0	01	0
	0	1	00	0
	1	1	00	0
	1	0	00	0
b = 01	0	0	00	0
	0	1	00	0
	1	1	11	0
	1	0	11	0
c = 11	0	0	00	0
	0	1	00	1
	1	1	00	0
	1	0	00	0

(b)

		AB = 00	01	11	10
	xy = 00	0	0	0	0
	01	0	0	0	0
A(t + 1)	11	0	1	0	0
	10	0	1	0	0
	00	1	0	0	0
	01	0	0	0	0
B(t + 1)	11	0	1	0	0
	10	0	1	0	0

Fig. 13.16. (a) Excitation table and (b) Karnaugh maps for the flip-flops in this trapping system.

It is straightforward to arrange this logic with RS or D flip-flops, e.g. $S_A = \overline{A}.B.x$, $R_A = \overline{S}_A$. Suppose, however, JK flip-flops are to be used instead. We need to use the characteristic equation of this flip-flop:

$$Q(t+1) = J.\overline{Q} + \overline{K}.Q.$$

Equation (13.3), for example, may be rewritten

$$B(t+1) = \overline{K}_B.B + J_B.\overline{B}$$

$$J_B = \overline{A}.\overline{x}.\overline{y} \tag{13.4}$$

$$\overline{K}_B = \overline{A}.x \qquad \text{or} \qquad K_B = A + \overline{x}. \tag{13.5}$$

Copyright © 2005 IOP Publishing Ltd.

Likewise, equation (13.2) may be rewritten

$$A(t+1) = J_A.\overline{A} + \overline{K}_A.A$$

$$J_A = B.x \tag{13.6}$$

$$\overline{K}_A = 0 \qquad \text{or} \qquad K_A = 1.$$

You may easily check using the Karnaugh maps that these equations give the correct transitions. For example, for the two entries in the top left hand corner of figure 13.16(a) with $A = B = x = y = 0$, B toggles ($J_B = K_B = 1$) and A resets ($K_A = 1$, $J_A = 0$).

There is one final point about constructing the logic of *JK* flip-flops not illustrated by this example. Sometimes it happens that the expression for $A(t+1)$, say, does not depend on A. For sake of example, the equation might be

$$A(t+1) = B.x. \tag{13.7}$$

This does not contain A or $\overline{A}$. How can this be written in the form of the characteristic equation $Q(t+1) = J_A.\overline{A} + \overline{K}_A.A$? The answer is simply to use the fact that $A + \overline{A} = 1$ and write equation (13.7) as

$$A(t+1) = B.x.(A + \overline{A}).$$

Then

$$J_A = B.x$$

$$\overline{K}_A = B.x \qquad \text{or} \qquad K_A = \overline{J}_A.$$

13.7 The Monostable

There are two further important circuits related to the flip-flop, but used in other applications. The first is a circuit which can be triggered to give an output pulse

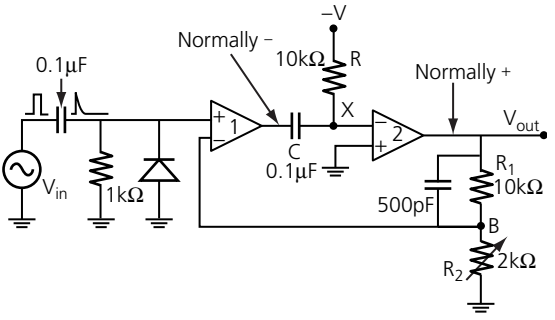


Fig. 13.17. The monostable

Copyright © 2005 IOP Publishing Ltd.

having a preset duration, independent of the input. Such a circuit is called a **monostable** or **univibrator**.

The monostable is like a flip-flop, but one of the two states is unstable and the second stable. This is achieved with the circuit of Figure 13.17. The bias voltage $-V$ keeps the output of amplifier 2 normally positive and that of amplifier 1 normally negative. However, a large enough positive input pulse to the + input of amplifier 1 drives its output positive. The output is transmitted through C, which cannot change its voltage instantly, and amplifier 2 changes state to a negative output; its output feeds back to amplifier 1, enhancing the switching action there.

Pulse length

Clearly this creates an unstable state. Subsequent waveforms are shown in figure 13.18. What happens is that C charges through R and when the voltage at X passes zero, amplifier 2 switches back to its normal stable state and takes amplifier 1 with it. In order to find the length of the output pulse, it is necessary to study the charging of C.

Suppose the output of amplifier 1 switches from $-V_{\downarrow}$ to $+V_{\uparrow}$. When the trigger pulse arrives and amplifier 1 switches, the capacitor drives X positive by an amount $V_{\uparrow} + V_{\downarrow}$. Then the subsequent voltage V_X is given by

$$V_X = -V + (V_{\uparrow} + V_{\downarrow})\mathrm{e}^{-t/CR};$$

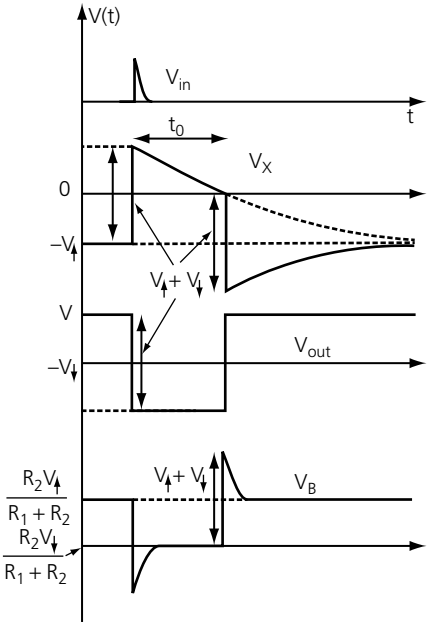


Fig. 13.18. Waveforms in the monostable.

Copyright © 2005 IOP Publishing Ltd.

V_X reaches zero when

$$V = (V_\uparrow + V_\downarrow)\mathrm{e}^{-t_0/CR}$$

$$t_0 = CR \ln_\mathrm{e}\left(\frac{V_\uparrow + V_\downarrow}{V}\right). \tag{13.8}$$

The length of the output pulse may be varied via V, R or C.

When the first amplifier switches back to its quiescent state, the capacitor C has to charge again to its normal level. Until this is complete, the circuit is not fully primed to receive another trigger pulse, though it will respond to a second trigger; however, because the capacitor is abnormally charged, the length of the output pulse will be altered. The time required for the capacitor to recharge is called its **settling time**. Various refinements are possible to return the capacitor to its steady state more quickly, e.g. by putting a diode across C to allow rapid charging in one direction.

Why are the capacitors included at all? Why not omit them and allow the opamps to respond directly to the input square pulse and the switching action of the second opamp? The answer is that there is always some capacitance C_{in} between the feedback terminals of the operational amplifier. The necessity to charge this capacitance slows down the risetime of the input to the amplifier. This may be remedied by using the **speed-up capacitor** C_1 which over-compensates any capacitance across the input of each amplifier and makes them switch promptly. However, time constants $C_1 R_1$ must be small compared with the interval between pulses.

13.8 The Pulse Generator

It is straightforward to make a circuit with no stable state. This is called the **astable**. It flips back and forth continuously between two states, logical 0 and 1. It has the important application of making a pulse generator. The circuit is shown schematically in figure 13.19(a) and the output waveforms in figure 13.19(b). The ratio τ_1/τ_2 is called the **mark-to-space ratio** for obvious reasons.

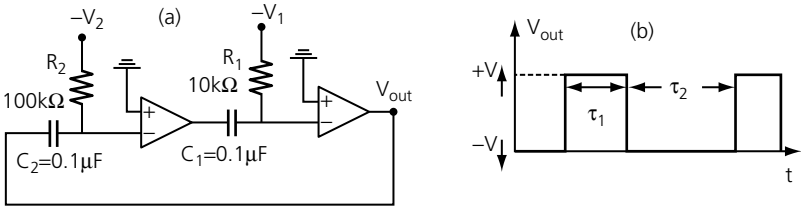


Fig. 13.19. (a) The astable, (b) its waveform.

Copyright © 2005 IOP Publishing Ltd.

13.9 Exercises

1. Two clocked *RS* flip-flops hold a 2-bit binary positive number AB. Design a circuit so that a clock pulse x converts 01 to 11, 11 to 01 and leaves 00 and 10 unchanged. (Ans: Figure 13.20.)

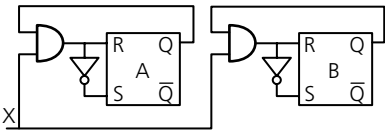


Fig. 13.20.

2. (RHBNC).

   Give the circuit of a counter of 3 *D* flip-flops which goes through the following sequence: ABC = 000, 100, 110, 111, 011, 001, 000. Draw the transition diagram of the unused states. (Ans: Figures 13.21 and 13.22.)

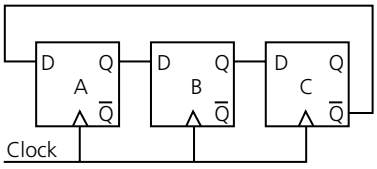


Fig. 13.21.

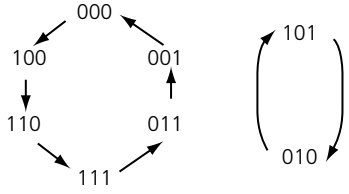


Fig. 13.22.

3*. (QMC). A five clock-cycle sequence generator is required to produce the waveforms for signals *A*, *B* and *C* as shown in figure 13.23. Design the sequential circuit to use three negative edge-triggered *T* flip-flops designated *FFA*, *FFB* and *FFC*; use don't care conditions. Show that if the input to *FFA* is $A + B.\underline{C}$, the circuit is self-starting. But if the input to *FFA* is $B.\overline{C}$, 100 → 111, 101 → 111 and 111 → 100, show the circuit is not self-starting. (Ans: figure 13.24.)

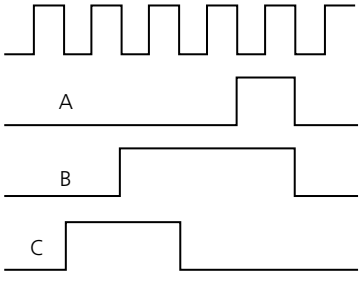


Fig. 13.23.

Copyright © 2005 IOP Publishing Ltd.

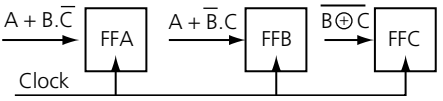


Fig. 13.24.

4. Show that Figure 13.25 using three *JK* negative edge-triggered flip-flops acts as a 3 digit binary counter. Sketch the following waveforms so as to show the relative times of the transitions in each waveform: (a) at the input, (b) at the output, (c) at the intermediate divide-by-and divide-by-4 outputs. [Note that this circuit is not typical of the use of JK flip-flops and is difficult to interpret in terms of the characteristic equation of the flip-flop]. (Ans: figure 13.26.)

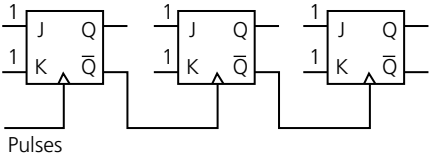


Fig. 13.25.

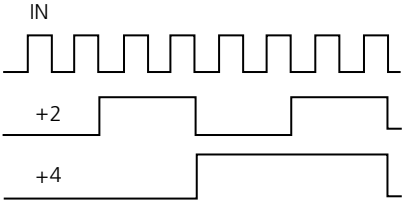


Fig. 13.26.

5. Design a counter using *JK* flip-flops which follows the sequence $00 \rightarrow 01 \rightarrow 10$ if an input x is 1 and the reverse sequence if $x = 0$. (Ans: Figure 13.27.)

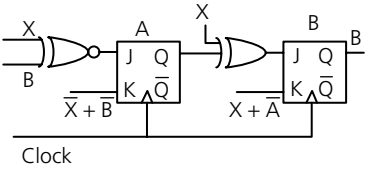


Fig. 13.27.

6. (RHBNC). Using *JK* flip-flops, design a synchronous code sequence generator circuit that gives the following code-words on successive clock

Copyright © 2005 IOP Publishing Ltd.

pulses: $ABC = 000, 001, 011, 010, 110, 111, 101, 100, 000$. (Ans: $J_A = B.\overline{C}$, $K_A = \overline{B}.\overline{C}$, $J_B = \overline{A}.C$, $K_B = A.C$, $J_C = A.B + \overline{A}.\overline{B}$, $K_C = \overline{J}_C$.)

7*. (QMC). Design a clocked sequential circuit using positive-edge triggered master-slave *JK* flip-flops to follow the state diagram shown in figure 13.28. The external input and the output are x and y respectively. Draw up the state table and derive the transition and excitation table required for the flip-flop inputs and the circuit output. What is the output? Show that one possible arrangement is $J_A = x.B$, $K_A = \overline{J}_A$, $J_B = 1$, $K_B = x$. Deduce what will happen if the unused states are accidentally entered and redraw the state diagram including the unused states. How can you ensure that the output $y = 1$ only occurs for the case shown in figure 13.28? (Ans: $y = \overline{x}.A.\overline{B}$; Figure 13.29; make $y = 0$ for unused states.)

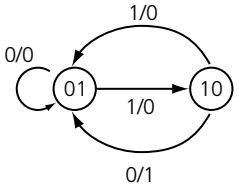


Fig. 13.28.

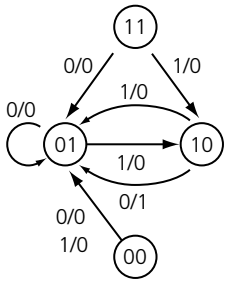


Fig. 13.29.

8. Using *JK* flip-flops, design a synchronous decade counter which will count from 0 to 9 in binary and then reset. (Ans: Figure 13.30.)

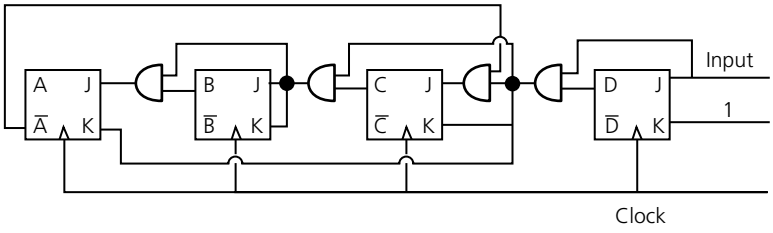


Fig. 13.30.

9**. (RHBNC). Draw the state transition diagram for the synchronous counter of figure 13.31. The flip-flops have no direct set or clear inputs. However, it is required to modify the circuit so that it can be reset on the next clock pulse when a given logic input $x = 1$. The circuit must also be self-starting and correcting by directing all states into the sequence which contains 0000. Show how to make the necessary modifications to the circuit. Show how the modulo of the counter may be multiplied by

Copyright © 2005 IOP Publishing Ltd.

two using only one extra flip-flop, while retaining synchronous operation. (Ans: 0000 → 1000 → 1100 → 1110 → 1111 → 0111 → 0011 → 0001 → 0000; 0010 → 1001 → 0100 → 1010 → 1101 → 0110 → 1011 → 0101 → 0010; $J_A = \overline{x}.\overline{D}$, $K_A = \overline{J}_A$, $J_B = \overline{x}.A$, $K_B = \overline{J}_B$, $J_C = \overline{x}.B$, $K_C = \overline{J}_C$, $J_D = \overline{x}.C$, $K_D = \overline{J}_D$; figure 13.32.)

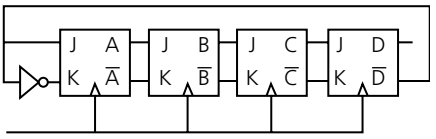


Fig. 13.31.

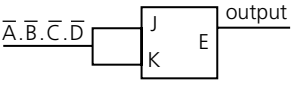


Fig. 13.32.

10. (QMC). (Lengthy.) Design a clocked sequential circuit using positive-edge triggered master-slave *JK* flip-flops to follow the state table given below.

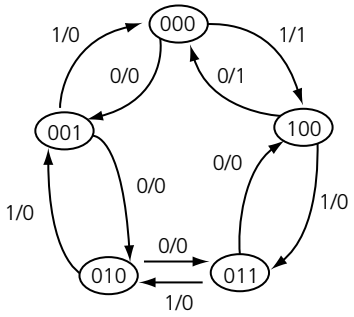


Fig. 13.33.

Present state	Next state $x = 0$ $x = 1$	Output, y $x = 0$ $x = 1$
a	b e	0 1
b	c a	0 0
c	d b	0 0
d	e c	0 0
e	a d	1 0

In this table x and y are the external input and output respectively. If the states a to e are encoded in binary as 000 to 100 consecutively, draw the encoded state diagram and derive the transition and excitation table required for the flip-flop inputs and the circuit output. Finally draw the logic diagram. (Ans: Figure 13.33; $J_A = x.\overline{B}.\overline{C} + \overline{x}.B.C$, $K_A = x + \overline{B} + \overline{C}$, $J_B = \overline{x}.C + x.A$, $K_B = \overline{x}.C + x.\overline{C}.\overline{A}$, $J_C = \overline{x}.\overline{A} + x.A + B$, $K_C = \overline{x} + \overline{A}$, $y = \overline{B}.\overline{C}.(x.\overline{A} + \overline{x}.A)$.)

Copyright © 2005 IOP Publishing Ltd.

14

Resonance and Ringing

14.1 Introduction

Resonance in mechanical systems is a familiar phenomenon. If you pluck the string of a violin it vibrates at a frequency depending on its length and tension. Likewise, rotating machinery is liable to resonate violently at certain speeds (e.g. a washing machine). We shall find that a circuit containing an inductance L and capacitance C also oscillates with a natural frequency $\omega_0 = 1/\sqrt{LC}$; over a range of frequencies close to this, large currents or voltages can be excited in the circuit.

This chapter will go through two explicit examples of resonance; a fair amount of algebraic manipulation is necessary, but is straightforward. These two cases may be regarded as worked examples of the ideas of impedance developed in Chapter 6. This work also serves as a useful example for Fourier analysis in the next chapter.

14.2 Resonance in a Series LCR Circuit

The classic example of a resonant circuit is shown in figure 14.1(a) and consists of L, C and R in series. As an exercise, this circuit will be discussed using (i) phasor diagrams, (ii) complex numbers.

From Kirchoff's voltage law,

$$V = V_R + V_C + V_L = RI + (1/C)\int I\ \mathrm{d}t + L\mathrm{d}I/\mathrm{d}t. \tag{14.1}$$

Suppose

$$I = I_0 \mathrm{e}^{\mathrm{j}\omega t} \tag{14.2}$$

$$V = V_0 \mathrm{e}^{\mathrm{j}(\omega t - \phi)} \tag{14.3}$$

so that ϕ is the angle by which the current *leads* the applied voltage. Then

$$V = (R + \frac{1}{\mathrm{j}\omega C} + \mathrm{j}\omega L)I. \tag{14.4}$$

Copyright © 2005 IOP Publishing Ltd.

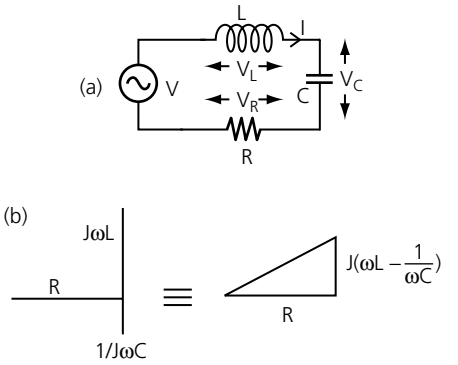


Fig. 14.1. (a) LCR series circuit and (b) its impedance diagram.

(i) From the impedance diagram, figure 14.1(b)

$$V_0 = I_0|Z| = I_0\{R^2 + (\omega L - 1/\omega C)^2\}^{1/2} \tag{14.5}$$

$$\text{Phase(V)} = \text{Phase(I)} - \phi = \text{Phase(I)} + \text{Phase(Z)}$$

$$\text{so} \qquad \phi = -\tan^{-1}(\omega L - 1/\omega C)/R. \tag{14.6}$$

(ii) The same conclusion can be reached algebraically using complex numbers from (14.2)–(14.4):

$$V_0 e^{-j\phi} e^{j\omega t} = [R + j(\omega L - 1/\omega C)] I_0 e^{j\omega t}. \tag{14.7}$$

Cancelling the factor $e^{j\omega t}$ and taking real and imaginary parts,

$$V_0 \cos\phi = R I_0 \tag{14.8}$$

$$-V_0 \sin\phi = (\omega L - 1/\omega C) I_0. \tag{14.9}$$

Squaring and adding (14.8) and (14.9),

$$V_0^2 = I_0^2\{(\omega L - 1/\omega C)^2 + R^2\}$$

in agreement with (14.5); and dividing (14.9) by (14.8):

$$\tan\phi = -(\omega L - 1/\omega C)/R$$

in agreement with (14.6).

Can you arrive at these same results using $V = V_0 \cos(\omega t - \phi)$ *and* $I = I_0 \cos\omega t$*?*

Copyright © 2005 IOP Publishing Ltd.

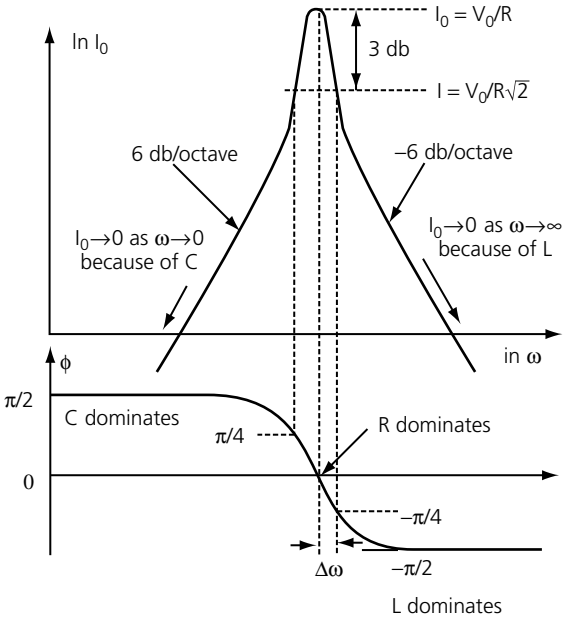


Fig. 14.2. Bode plot for the LCR series circuit.

Bode plot

The fundamental results of these alternative approaches are contained in equations (14.5) and (14.6) for I_0 and ϕ. At this point, it is however necessary to mention a source of confusion over the sign of the phase. In mechanics and in atomic and nuclear physics it is conventional to plot instead the angle $\psi = -\phi$; this would correspond here to the angle by which current *lags* voltage; i.e. if $V = V_0 \exp \mathrm{j}\omega t$, $I = I_0 \exp \mathrm{j}(\omega t - \psi)$. On figure 14.2, the convention used in electrical engineering is followed and ϕ is displayed; the plot is made for a small value of R giving a narrow peak. The curve for I_0 is called a **resonance** curve. The frequency ω_0 of the peak is given by

$$\omega_0 L = 1/\omega_0 C$$

$$\boxed{\omega_0 = 1/\sqrt{LC}.} \tag{14.10}$$

At this frequency, the reactances of L and C cancel, so $\phi = 0$ and I_0 peaks at the value V_0/R. As $\omega \to 0$, C dominates; as $\omega \to \infty$, L dominates.

Resonance width

The circuit may be used as a narrow-band filter (e.g. in tuning a radio) since it only transmits significant current near $\omega = \omega_0$. It is of interest to find a

Copyright © 2005 IOP Publishing Ltd.

quantitative measure of the width of the peak. It turns out that this is related to power considerations. The average power dissipated is:

$$P = \frac{1}{2} I_0^2 Re(Z) = \frac{\frac{1}{2} V_0^2 R}{(\omega L - 1/\omega C)^2 + R^2}. \tag{14.11}$$

On either side of the resonant peak, the power falls to half its maximum value when $(\omega L - 1/\omega C) = \pm R$, i.e. $\tan\phi = \pm 1$ or $\phi = \pm\pi/4$. At this point the current falls from its maximum value by a factor $1/\sqrt{2}$. The half-width $\Delta\omega$ of the peak is obtained by setting $\omega = \omega_0 \pm \Delta\omega$. Then

$$(\omega_0 \pm \Delta\omega)L - \frac{1}{(\omega_0 \pm \Delta\omega)C} = \pm R. \tag{14.12}$$

To simplify this expression, the binomial expansion is used to make the approximation:

$$(\omega_0 \pm \Delta\omega)^{-1} = \frac{1}{\omega_0}\left(1 \pm \frac{\Delta\omega}{\omega_0}\right)^{-1} \simeq \frac{1}{\omega_0}\left(1 \mp \frac{\Delta\omega}{\omega_0}\right).$$

Then equation (14.12) reads

$$(\omega_0 \pm \Delta\omega)L - \frac{1}{\omega_0 C}\left(1 \mp \frac{\Delta\omega}{\omega_0}\right) = \pm R$$

and using (14.10)

$$R = \Delta\omega(L + 1/\omega_0^2 C) = \Delta\omega(L + L)$$

so finally

$$\Delta\omega = R/2L. \tag{14.13}$$

The width of the peak is proportional to R and hence to the power dissipated. The resonance peak is narrow for small R, and its height is V_0/R. The product of height $\times$ width is independent of R. A narrow, high resonance can be changed to a wide, low one by increasing R.

Q value

A popular measure of the sharpness of the resonance curve is the **quality factor** Q. For resonances in general this is defined by

$$Q = 2\pi \times \frac{\text{maximum energy stored}}{\text{energy dissipated per cycle}}. \tag{14.14}$$

For the LCR series circuit,

$$\boxed{Q = \frac{2\pi \times \frac{1}{2} L I_0^2}{\frac{1}{2} I_0^2 R/f_0} = \frac{\omega_0 L}{R} = \frac{\omega_0}{\text{full width}}.} \tag{14.14a}$$

Copyright © 2005 IOP Publishing Ltd.

This latter form is easy to remember. The same result will emerge for parallel resonance discussed later. A high Q circuit gives a narrow resonance and small bandwidth.

The treatment here ignores the coupling of the circuit to electromagnetic waves. Any oscillatory circuit like this acts as an antenna and radiates radio waves. If the radiation is significant, it adds to the width of the curve (**power broadening**). To take account of this, an equivalent series resistance called **radiation resistance** should really be inserted into the circuit diagram:

$$R_{\text{radiation}} = 2 \times \text{Power radiated}/\text{I}_0^2.$$

The voltages across C and L are obtained from $I/\text{j}\omega C$ and $(\text{j}\omega L)I$. At resonance, they have magnitudes $V_0/\omega_0 CR$ and $V_0\omega L/R$ which may be large. For example, suppose $R = 10\,\Omega$, $C = 0.1\,\mu$F and $L = 4$ mH. Then $\omega_0 = 5 \times 10^4$ rad/s and $V_C = V_L = 20V_0$, i.e. much larger than the applied voltage. On the phasor diagram, figure 14.1(b), V_C and V_L are individually large compared with V_R, but on resonance cancel exactly. Off resonance, the difference between V_C and V_L swamps V_R if the resonance is narrow. Because V_C and V_L are individually large, this type of resonance is called **voltage resonance**. The narrower the resonance, the larger these voltages are; for a very narrow resonance, there is the danger that they actually damage components. What is happening is that a small applied voltage $V_0\text{e}^{\text{j}\omega t}$ is exciting large resonant voltages in C and L. The next sections will examine more closely just how this comes about.

Incidentally, the name voltage resonance is potentially confusing, since the magnitude of the current I_0 also peaks at $\omega = \omega_0$; this is because voltages V_L and V_C have opposite signs, so they cancel, leaving the full applied voltage across R.

14.3 Transient in a CL Circuit

The resonant frequency $\omega_0 = 1/\sqrt{LC}$ depends only on C and L. This is a clue that these two elements should be considered alone, figure 14.3. Transient oscillations in this circuit will be discussed. The result which emerges is that the current performs simple harmonic oscillations. There is no damping because there is no resistance to dissipate energy.

For this circuit,

$$V = L\text{d}I/\text{d}t + (I/C)\int I\,\text{d}t$$

$$\text{or} \qquad C\ \text{d}V/\text{d}t = CL\text{d}^2I/\text{d}t^2 + I.$$

Suppose the switch is closed at $t = 0$. Except at this instant, the left-hand side is zero, resulting in the equation

$$\text{d}^2I/\text{d}t^2 + I/CL = 0. \tag{14.15}$$

Copyright © 2005 IOP Publishing Ltd.

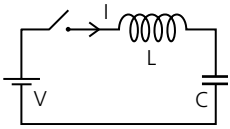


Fig. 14.3. CL resonant circuit.

This is the equation of **simple harmonic motion**

$$I = I_0 \cos(\omega_0 t + \phi_1) \tag{14.16}$$

as is now demonstrated by explicit differentiation and back-substitution into (14.15):

$$\mathrm{d}I/\mathrm{d}t = -\omega_0 I_0 \sin(\omega_0 t + \phi_1)$$

$$\mathrm{d}^2 I/\mathrm{d}t^2 = -\omega_0^2 I_0 \cos(\omega_0 t + \phi_1) = -\omega_0^2 I.$$

This satisfies equation (14.15) with

$$\omega_0^2 = 1/CL.$$

So the current executes natural oscillations at angular frequency $\omega_0 = 1/\sqrt{CL}$. Energy is stored alternately as kinetic energy $\frac{1}{2}I^2 L$ in the form of current and as potential energy $\frac{1}{2}Q^2/C$ in the form of charge on the capacitor. As long as there is no resistance in the circuit, this oscillation continues indefinitely. The resonance in the LCR circuit is clearly associated with this oscillation; the AC voltage excites large oscillations in L and C if the applied frequency is close to that of natural oscillations.

14.4 Transient in the Series LCR Circuit

After this preliminary, let us return to the LCR series circuit of figure 14.4. In this circuit the resistance dissipates energy, so when the switch is closed the current executes the damped oscillations shown in figure 14.5. This is called **ringing** and is a common result of a pulse in an electrical circuit.

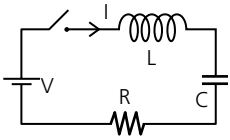


Fig. 14.4. A switched LCR series circuit.

Copyright © 2005 IOP Publishing Ltd.

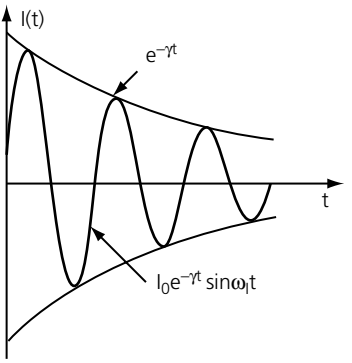


Fig. 14.5. Damped simple harmonic motion.

The algebra is only a minor extension of the previous case. With the addition of resistance R,

$$V = L\mathrm{d}I/\mathrm{d}t + RI + (1/C)\int I\,\mathrm{d}t$$

$$\text{or} \qquad C\frac{\mathrm{d}V}{\mathrm{d}t} = CL\frac{\mathrm{d}^2I}{\mathrm{d}t^2} + CR\frac{\mathrm{d}I}{\mathrm{d}t} + I \tag{14.17}$$

and, except at the instant when the switch is closed, the left-hand side is zero. The equation is rewritten in the standard form

$$\boxed{\mathrm{d}^2I/\mathrm{d}t^2 + 2\gamma\mathrm{d}I/\mathrm{d}t + \omega_0^2 I = 0} \tag{14.18}$$

where

$$\gamma = R/2L \tag{14.19}$$

$$\omega_0^2 = 1/CL. \tag{14.20}$$

This equation is easily solved by the trial substitution:

$$I = I_0 \mathrm{e}^{st+\phi} \tag{14.21}$$

which gives

$$(s^2 + 2\gamma s + \omega_0^2)I = 0$$

or

$$s = -\gamma \pm (\gamma^2 - \omega_0^2)^{1/2} \tag{14.22}$$

$$s = -\gamma \pm \mathrm{j}(\omega_0^2 - \gamma^2)^{1/2}. \tag{14.22a}$$

The latter expression describes oscillations if the quantity

$$\omega_1^2 = \omega_0^2 - \gamma^2 \tag{14.23}$$

is positive, i.e. for small R. As $R \to 0$, $\gamma \to 0$ and $\omega_1 \to \omega_0$ and the results of the previous section reappear.

Copyright © 2005 IOP Publishing Ltd.

Damped SHM

The two alternative values of s give the general solution

$$I = \mathrm{e}^{-\gamma t}\{A\mathrm{e}^{\mathrm{j}\omega_1 t} + B\mathrm{e}^{-\mathrm{j}\omega_1 t}\}. \tag{14.24}$$

For I to be real, A and B must be complex with $B = A^*$. Then

$$I = 2\mathrm{e}^{-\gamma t}(Re A \cos\omega_1 t - Im A \sin\omega_1 t)$$

which may be written in the alternative forms

$$I = \mathrm{e}^{-\gamma t}\{A' \cos\omega_1 t + B' \sin\omega_1 t\} \tag{14.24a}$$

$$\text{or} \qquad I = I_0\mathrm{e}^{-\gamma t}\sin(\omega_1 t + \phi) \tag{14.24b}$$

where

$$I_0 \cos\phi = B' \qquad I_0 \sin\phi = A'.$$

The two constants A' and B' are to be fixed by the initial conditions. If $I = 0$ when $t = 0$, then $\phi = 0$ and

$$I = I_0\mathrm{e}^{-\gamma t}\sin\omega_1 t. \tag{14.24c}$$

This equation describes damped simple harmonic motion. The envelope of the oscillation in figure 14.5 is exponentially damped by the factor $\mathrm{e}^{-\gamma t}$. As expected, the damping constant $\gamma = 2R/L$ is proportional to R. The damping reduces the frequency of natural oscillaltion from ω_0 to $\omega_1 = \sqrt{\omega_0^2 - \gamma^2}$.

Can you see the relation between I_0 and V_0 in (14.24c). Clue: what is $\mathrm{d}I/\mathrm{d}t$ just after the switch is closed?

There is an intimate relation between the transient behaviour of figure 14.5 and the resonance curve of figure 14.2. An AC signal close to frequency ω_1 finds it easy to drive the LCR circuit. In fact, the driving voltage is supplying energy at exactly the rate at which the circuit dissipates it. The oscillations are largest for small damping. The damping constant $\gamma = R/2L$ of the transient is equal to the half-width of the resonance curve. This is no accident: Chapter 15 reveals that it is a general consequence of Fourier's theorem.

Overdamping

If the damping is large, ω_1^2 may become negative, in which case there is no oscillation and the values of s given by (14.22) are appropriate. Then

$$I = I_0\mathrm{e}^{-\gamma t}(A\mathrm{e}^{-\Omega t} + B\mathrm{e}^{\Omega t}) \tag{14.25}$$

$$\text{where} \qquad \Omega = (\gamma^2 - \omega_0^2)^{1/2}. \tag{14.26}$$

The current is now the sum of two falling exponentials, figure 14.6. The circuit is said to be **overdamped**. At large enough times, the more slowly falling exponential dominates.

Copyright © 2005 IOP Publishing Ltd.

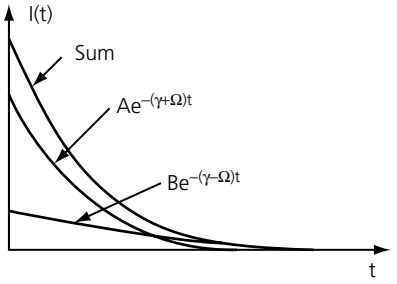


Fig. 14.6. Overdamped response.

Critical damping

In electronics, and particularly in servo-systems, the circuit should generally follow the input as faithfully as possible. A step in the input should be followed with a step in the output. Overshooting and ringing, figure 14.7, are to be avoided. Conversely, we want to avoid the sluggish response which is a consequence of overdamping, figure 14.6. The optimum response is **critically damped**, corresponding to $\gamma = \omega_0$ and $\Omega = 0$.

14.5 Parallel LCR

A second very common circuit displaying resonance is the parallel LC arrangement shown in figure 14.8. It is called a **tuned circuit**. We shall find that V_{out} follows a resonance curve very similar to figure 14.2; and if a voltage step is applied, the current executes damped oscillations very similar to the series LCR circuit. The equations can be manipulated into a form identical to those of the previous sections with minor changes of parameters.

When V and R are replaced with their Norton equivalent circuit, as in figure 14.9, the resistance appears in parallel with L and C. If there is a resonant current in L and C, some current is diverted to R and dissipates energy, damping the oscillation.

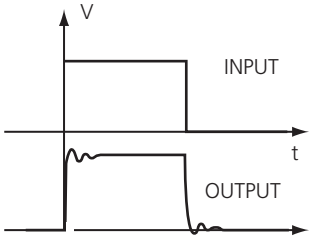


Fig. 14.7. Ringing response to a step function input.

Copyright © 2005 IOP Publishing Ltd.

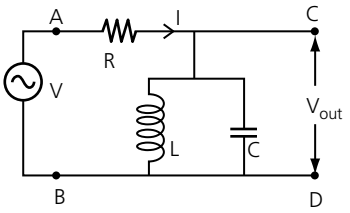


Fig. 14.8. Parallel LC circuit.

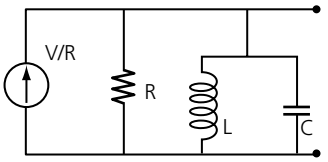


Fig. 14.9. Norton equivalent of figure 14.8.

Consider first the response to an AC voltage. The LC arrangement has impedance Z_{CL} where

$$1/Z_{CL} = \mathrm{j}\omega C + 1/\mathrm{j}\omega L = (1 - \omega^2 CL)/\mathrm{j}\omega L.$$

From figure 14.8, $V = IZ$ with

$$Z = R + \mathrm{j}\omega L/(1 - \omega^2 CL).$$

The phasor diagram is shown in figure 14.10. If $V = V_0 \exp \mathrm{j}(\omega t - \phi)$ and $I = I_0 \exp \mathrm{j}\omega t$,

$$I_0 = V_0 / \left[R^2 + \left(\frac{\omega L}{1 - \omega^2 CL} \right)^2 \right]^{1/2} \tag{14.27}$$

$$\tan \phi = \omega L/\{R(\omega^2 CL - 1)\}. \tag{14.28}$$

There is a resonant peak in V_{out} of figure 14.8 very similar to figure 14.2, centred at $\omega = 1/\sqrt{LC}$, where the impedances $\mathrm{j}\omega L$ and $1/\mathrm{j}\omega C$ balance. At resonance, $Z_{CL} \to \infty$ and I_0 drops to 0; however, large balancing AC currents I_1 and I_2 flow through L and C individually. For this reason, the resonance is called **current resonance**. On resonance, $I_0 \to 0$, so there is no voltage drop across R and the full applied voltage $V_0 \exp \mathrm{j}\omega t$ appears across the tuned circuit. Off resonance, this voltage falls by a factor $1/\sqrt{2}$ when $\tan \phi = \pm 1$, i.e.

$$\pm \omega L/R = \omega^2 CL - 1.$$

Using the same approximations as in the LCR case, this occurs for $\omega = \omega_0 \pm \Delta\omega$ where

$$\Delta\omega \simeq 1/2CR. \tag{14.29}$$

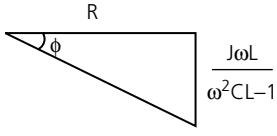


Fig. 14.10. Phasor diagram for figure 14.8.

Copyright © 2005 IOP Publishing Ltd.

On resonance, $|I_L| = V_0/\omega L$ and $|I_C| = \omega C V_0$. The Q of the circuit is

$$A = \frac{2\pi \times \frac{1}{2} I_L^2 L}{(\frac{1}{2} V_0^2/R)/f_0} = \omega_0 \frac{R}{\omega_0^2 L} = \omega_0 RC = \frac{\omega_0}{2\Delta\omega} = \frac{\omega_0}{\text{full width}}$$

as for the LCR series circuit.

Coil resistance

In practice, the inductor will have a small resistance r. So far this has been neglected in the interests of keeping algebra to a minimum. There is a useful trick which allows it to be included approximately if it is small. The admittance of the parallel CL combination is

$$Y_{CL} = \mathrm{j}\omega C + \frac{1}{\mathrm{j}\omega L + r} = \mathrm{j}\omega C + \frac{1}{\mathrm{j}\omega L(1 + r/\mathrm{j}\omega L)}. \tag{14.30}$$

Using the binomial theorem to expand $(1 + r/\mathrm{j}\omega L)$,

$$Y_{CL} \simeq \mathrm{j}\omega C + \frac{1 - r/\mathrm{j}\omega L}{\mathrm{j}\omega L} = \mathrm{j}\omega C + \frac{1}{\mathrm{j}\omega L} + \frac{r}{\omega^2 L^2}. \tag{14.30a}$$

This is the same admittance as is given by a parallel combination of C with a pure inductance L and a resistor $R' = \omega^2 L^2/r = L/Cr$. The resistor R' may be combined in parallel with R of figure 14.9 to make R'' and thereafter the previous algebra goes through with R'' replacing R. On resonance, the impedance of the tuned circuit is R', and is proportional to $1/r$. Because of R', the impedance of the tuned circuit is not infinite on resonance, so V_{out} of figure 14.8 is less than V. This allows an easy measurement of R'.

Considering energy dissipation, why does the series LCR circuit give a broad peak if R is large while the parallel LCR circuit gives a narrow peak?

Any circuit contains a small amount of stray capacitance and stray inductance. Why don't all circuits resonate at high frequency?

Transients

It will now be demonstrated that transients in the parallel LC circuit of figure 14.11 obey the same equation for damped simple harmonic motion as for the series LCR

Copyright © 2005 IOP Publishing Ltd.

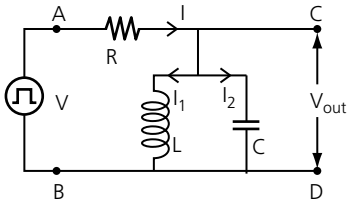


Fig. 14.11. Square pulses applied to the parallel CL circuit.

circuit, except for a different value of γ. Because there is the same voltage across L and C

$$L\mathrm{d}I_1/\mathrm{d}t = (1/C)\int I_2\mathrm{d}t$$

$$\text{so} \qquad I_2 = CL\mathrm{d}^2I_1/\mathrm{d}t^2.$$

Then

$$V = RI + L\mathrm{d}I_1/\mathrm{d}t = R(I_1 + I_2) + L\mathrm{d}I_1/\mathrm{d}t$$

$$\text{or} \qquad V/CLR = d^2I_1/\mathrm{d}t^2 + (1/CR)\mathrm{d}I_1/\mathrm{d}t + \omega_0^2 I_1. \tag{14.31}$$

This is the same equation as (14.18) except that (a) $\gamma = 1/2CR$, (b) the constant term on the left-hand side adds a DC current V/R; it is obvious that this must be so after the transient oscillations have died away. Notice that again the damping constant γ is equal to the half-width $\Delta\omega$ of the resonance curve, given by equation (14.29).

It is instructive to demonstrate both resonance and transient oscillations for yourself on the oscilloscope using the circuits of figures 14.11 and 14.8. The input signal V_{AB} may be displayed on one trace and the signal V_{CD} across the tuned circuit on the second. Suitable circuit parameters are $R = 10^5\ \Omega$, $C = 0.01\ \mu$F and $L = 4$ mH giving $\omega_0 \simeq 1.6 \times 10^5$ rad/s or $f_0 \simeq 25$ kHz. If you omit C, the circuit will resonate at a higher frequency due to stray capacitance across the coil. You will probably find a discrepancy with the calculated width of the resonance curve (and R'). At high frequencies, current is concentrated in the surface of the wire by the skin effect (see textbooks on electromagnetic theory) and the resistance r of the coil increases significantly above its DC value.

If you switch to square waves and reduce the frequency of the generator to ~100 Hz, you can examine the transients. By varying the resistance R or by putting a variable 4 kΩ resistor in parallel with the tuned circuit, you can vary the damping and demonstrate that oscillations disappear when $\gamma = \omega_0$. It is also interesting to insert a soft iron core into the inductor. This increases the value of L, therefore reducing ω_0 and increasing the impedance R' of the tuned circuit on resonance, hence the magnitude of the output signal.

Copyright © 2005 IOP Publishing Ltd.

14.6* Poles and Zeros

As you vary the frequency of the generator in figure 14.8 through resonance, the output signal peaks and the phase ϕ moves rapidly through 90°. The algebra will now be recast slightly so as to display better the origin of this rapid phase variation. An understanding of this point will be helpful in a wide range of advanced phenomena where damped oscillations occur, e.g. in servosystems.

If $V = V_0 e^{j\omega t}$, equation (14.17) gives

$$F e^{j\omega t} = d^2I/dt^2 + 2\gamma\, dI/dt + \omega_0^2 I \tag{14.32}$$

where $F = j\omega V_0/L$. Equation (14.31) gives a similar result except for a slight change in F. Suppose $I = I_1 e^{j\omega t}$ where I_1 is complex and includes the phase dependence. Substituting this into (14.32),

$$F = I_1(-\omega^2 + 2j\gamma\omega + \omega_0^2).$$

The right-hand side can be factorised into $-I_1(\omega - \omega_A)(\omega - \omega_B)$ where $j\omega_A = s_A$ and $j\omega_B = s_B$ are the two solutions (14.22). After these manipulations,

$$I_1 = \frac{-F}{(\omega - \omega_1 - j\gamma)(\omega + \omega_1 - j\gamma)}. \tag{14.33}$$

The magnitude of I_1 peaks around the resonant frequency $\omega = \omega_1$ because the first term in the denominator is small. The second term does not vary rapidly and can be approximated in the vicinity of the resonance by $(\omega + \omega_1 - j\gamma) \simeq 2\omega_1$. Then

$$I_1 \simeq \frac{-F/2\omega_1}{\omega - \omega_1 - j\gamma}. \tag{14.34}$$

The denominator goes to zero when $\omega = \omega_1 + j\gamma$. This is called a **pole** or **singularity** of I_1. The value of ω at the pole is complex, so it is not a physically realisable frequency. Physically ω can vary only along the real axis of figure 14.12(a) from $\omega = 0$ to ∞. However, as a piece of mathematics the function $|I_1|$ can be plotted for both real and imaginary values of ω. (This is called the complex ω plane). What happens, figure 14.12(b), is that $I_1 \to \infty$ for $\omega = \omega_1 + j\gamma$. The function has a surface shaped like a volcano centred at the pole. In the vicinity of this (complex) frequency, the behaviour of I_1 is dominated by this singularity. What is measured physically is the slice along the real ω axis. The resonance is narrow and high if the pole lies close to the real axis, i.e. if $j\gamma$ is small. Near the resonance, the complex value of I_1 depends on 1/(complex distance from the pole). This gives instantly the two decisive features of a resonance: a peak in the magnitude and a rapid phase variation. The phase variation goes hand in hand with the resonant peak and (except in contrived situations) one cannot arise without the other.

Copyright © 2005 IOP Publishing Ltd.

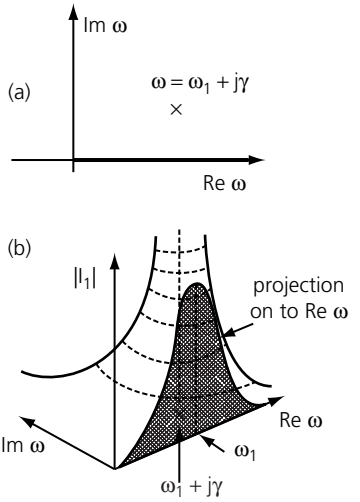


Fig. 14.12. (a) Location of the pole in the complex ω plane, (b) 3-D plot of I_1 for complex values of ω.

For parallel resonance, the current goes to zero as $\omega \to \omega_0$. For series resonance, the current peaks at $\omega = \omega_0'$. It may be shown that the impedance of *any* circuit may be factorised into the form

$$Z = \frac{(\omega - \omega_0)(\omega - \omega_1)\ldots(\omega - \omega_n)}{(\omega - \omega_A)(\omega - \omega_B)\ldots(\omega - \omega_N)},$$

where values of ω_0–ω_n and ω_A–ω_N are complex. Voltage resonances occur for values of ω near the zeros ω_0–ω_n and current resonances for ω close to the poles ω_A–ω_N.

14.7 Exercises

1. What is the origin of resonance in a series LCR circuit? At what frequency does the resonance appear? How is this related to current resonance when capacitor and inductor are in parallel and the combination is in series with a resistor? For each case of resonance, what dictates the width of the resonance? Do small resistors produce wide or narrow resonances?

2. (QMC). In figure 14.13, S is a constant voltage generator of variable frequency and zero internal impedance. For what range of frequencies can this circuit be tuned by placing a variable capacitance across the terminals AB? Sketch a curve showing the amplitude of the voltage across the $1\,\Omega$ resistor as a function of frequency if a $1\ \mu$F capacitor is connected across AB. What is the resonant frequency and the half-width $\Delta\omega$ of the resonance curve? (Ans: $\omega > 2 \times 10^4$ rad/s; $\omega_0 = 3.7 \times 10^4$ rad/s; $\Delta\omega = 500$ rad/s.)

Copyright © 2005 IOP Publishing Ltd.

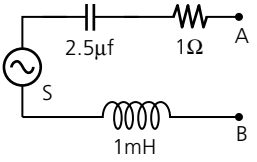


Fig. 14.13.

3. Define the Q of a series LCR circuit. If $Q = 100$, $R = 50\ \Omega$ and the resonant frequency ω_0 is 20×10^3 rad/s, what are the values of L, C and the half-width $\Delta\omega$ of the resonance? (Ans: $\Delta\omega = 10^2$ rad/s, $L = 0.25$ H, $C = 0.01\ \mu$F.)

4. (QMC). Find the current and the voltage across each element of figure 14.14 at resonance and sketch how current varies with ω over the range 0.5 to 1.5 times the resonant frequency. (Ans: 10 mA, 5 V, 5 V, $\omega_0 = 2 \times 10^4$ rad/s, $\Delta\omega = 2 \times 10^3$ rad/s.)

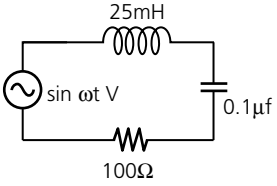


Fig. 14.14.

5. A parallel LCR circuit has a lossy inductor with resistance r. The circuit is driven by an AC voltage source of angular frequency ω. Show that at resonance (when voltage and current from the source are in phase), $\omega^2 = (1/LC) - r^2/L^2$.

6*. In figure 14.15, the capacitor is charged to voltage V_0 with the switch open. What is this voltage and $\mathrm{d}V/\mathrm{d}t$ immediately after the switch is closed at time $t = 0$? Show that subsequent oscillations obey the equation $CL\mathrm{d}^2V/\mathrm{d}t^2 + L/R\mathrm{d}V/\mathrm{d}t + V = 0$. Show that the solution to this equation is $V(t) = V_0 \exp -(t/2CR)\{\cos\omega_1 t - (1/2\omega_1 CR)\sin\omega_1 t\}$, where $\omega_1^2 = \{(1/CL) - (1/4C^2R^2)\}^{1/2}$. What are the frequency and decay time of the oscillations if $L = 4$ mH, $C = 0.5\ \mu$F, and $R = 20\ \mathrm{k}\Omega$? (Ans: $V = V_0$, $(\mathrm{d}V/\mathrm{d}t)_t = 0 = -V_0/RC$; $\omega_1 = 2.24 \times 10^4$ rad/s, $\tau = 0.02$ s.)

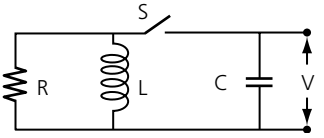


Fig. 14.15.

Copyright © 2005 IOP Publishing Ltd.

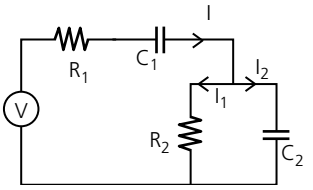


Fig. 14.16.

7*. For the circuit of figure 14.16, show that I_2 obeys the differential equation

$$C_1C_2R_2\frac{\mathrm{d}^2V}{\mathrm{d}t^2} = C_1C_2R_1R_2\frac{\mathrm{d}^2I_2}{\mathrm{d}t^2} + (C_1R_1 + C_1R_2 + C_2R_2)\frac{\mathrm{d}I_2}{\mathrm{d}t} + I_2.$$

Show that this circuit does not show ringing (i.e. natural oscillations) for any values of C_1, C_2, R_1 and R_2.

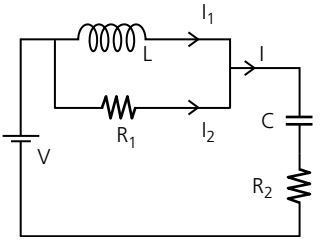


Fig. 14.17.

8*. Find the equation governing transients in I_1 in the circuit of figure 14.17. If $\tau_1 = L/R_1$ and $\tau_2 = CR_2$, show that $(CL + \tau_1\tau_2)\mathrm{d}^2I_1/\mathrm{d}t^2 + (\tau_1 + \tau_2)\mathrm{d}I_1/\mathrm{d}t + I_1 = 0$. Show that the circuit rings if $4CL > (\tau_1 - \tau_2)^2$.

Copyright © 2005 IOP Publishing Ltd.

15

Fourier's Theorem

15.1 Introduction

Up to here, we have dealt largely with two special cases:

(a) DC, where the applied voltage is constant,
(b) AC of a single frequency: $V_{\text{applied}} = V_0 \cos \omega t$.

These methods need to be extended to deal with the possibility of several frequencies being superimposed and with voltages of any shape $V(t)$.

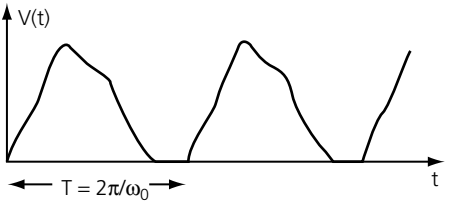


Fig. 15.1. A repeating waveform.

Suppose initially that an applied voltage is repeated at a fixed frequency ω_0, as in figure 15.1, but the waveform is complicated. A musical sound is an example. Different musical instruments playing the same note produce waveforms of different shapes, but the same repetition rate. Fourier's theorem states that $V(t)$ can be expressed as a series of terms at angular frequency ω_0 and multiples of it (harmonics):

$$V(t) = a_0 + a_1 \sin(\omega_0 t + \phi_1) + a_2 \sin(2\omega_0 t + \phi_2) + \ldots \tag{15.1}$$

$$= \sum_{n=0}^{\infty} a_n \sin(n\omega_0 t + \phi_n).$$

Copyright © 2005 IOP Publishing Ltd.

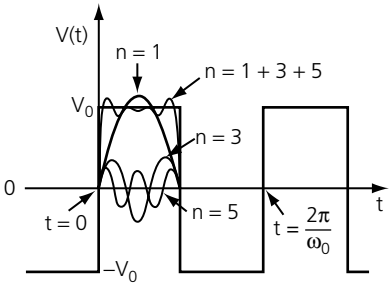


Fig. 15.2. A square wave and its Fourier components.

Consider an example where $V(t)$ is a square wave, figure 15.2. Then the Fourier series is

$$V(t) = \frac{4V_0}{\pi}(\sin\omega_0 t + \frac{1}{3}\sin 3\omega_0 t + \frac{1}{5}\sin 5\omega_0 t + \ldots). \tag{15.2}$$

The way successive harmonics build up the square wave is illustrated in figure 15.2. The fundamental is a sine wave, but as more harmonics are added, the waveform approaches the shape of the square wave progressively; see for example $n = 1 + 3 + 5$.

Suppose this square wave is applied to a circuit with impedance $Z(\omega)$. The current $I(\omega)$ may be found for each separate harmonic from $I(\omega) = V(\omega)/Z(\omega)$; then $I(t)$ may be built up by adding these components using the superposition principle. The important conclusion is that the general case for $V(t)$ can be solved with a superposition of simple AC solutions with suitable coefficients. Computer package are readily available to calculate terms of the Fourier series, then trace the behaviour of each one through circuits and add them up again to form the resulting complicated output. MATLAB and SIMULINK are commercial examples of such packages. Here we shall follow the principles.

15.2 A Square Wave applied to a CR Filter

Suppose a square wave is applied to the input of the filter circuit of figure 15.3(a) at a frequency ω_0 well below the cut-off frequency. The fundamental and low harmonics will be transmitted faithfully, but high harmonics above the cut-off are attenuated (and modified in phase). Consequently, the output will have rounded shoulders, figure 15.3b.

An oscilloscope itself has some input capacitance C_{in} and therefore a bandwidth limited to $\omega \leq 1/C_{in}R_{in}$. Even if perfectly square voltage pulses could be applied to its input, the time constant observed on the screen would be limited to

$$\tau = C_{in}R_{in}. \tag{15.3}$$

Copyright © 2005 IOP Publishing Ltd.

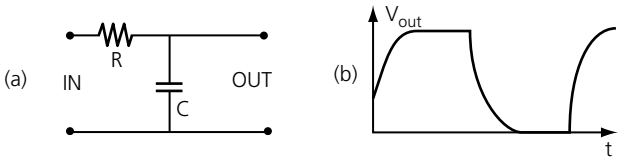


Fig. 15.3. (a) A low pass filter, (b) $V_{out}(t)$.

Conversely, a high pass filter cuts off low frequencies. If the cut-off frequency is less than that of the square waves, only DC is rejected; the DC term arises from $n = 0$ in equation (15.1).

If the waveform shown in figure 15.3(b) is applied to the terminals of figure 15.3(a) labelled OUT, will a square wave be observed at the terminals labelled IN?

15.3 How to Find Fourier Coefficients

Another way of writing equation (15.1) is

$$V(t) = b_0 + b_1 \cos\omega_0 t + b_2 \cos 2\omega_0 t + \cdots + a_1 \sin\omega_0 t + a_2 \sin 2\omega_0 t + \cdots$$
$$= \sum_{n=0}^{\infty} [a_n \sin(n\omega_0 t) + b_n \cos(n\omega_0 t)] .$$

If both sides are multiplied by $\sin(m\omega_0 t)$ and integrated over one complete cycle of the waveform from $t = 0$ to $T = 2\pi/\omega_0$,

$$\int_0^{2\pi/\omega_0} V(t) \sin(m\omega_0 t)\, \mathrm{d}t = \sum_{n=0}^{\infty} \int_0^{2\pi/\omega_0} \{a_n \sin(n\omega_0 t) + b_n \cos(n\omega_0 t)\} \sin(m\omega_0 t)\, \mathrm{d}t$$

it turns out that all but one of the terms on the right-hand side is zero. To see this, remember that

$$2\sin(n\omega_0 t)\sin(m\omega_0 t) = \cos\{(n-m)\omega_0 t\} - \cos\{(n+m)\omega_0 t\}$$
$$2\cos(n\omega_0 t)\sin(m\omega_0 t) = \sin\{(n+m)\omega_0 t\} - \sin\{(n-m)\omega_0 t\}.$$

When the cosines and sines on the right-hand side are integrated over a complete cycle they give zero, except for the cosine term with $n - m = 0$. Thus

$$a_n = (\omega_0/\pi) \int_0^T V(t) \sin(n\omega_0 t)\, \mathrm{d}t \tag{15.4}$$

Copyright © 2005 IOP Publishing Ltd.

or

$$\boxed{a_n = (1/\pi)\int_0^{2\pi} V(t)\sin(nx)\,\mathrm{d}x} \qquad (15.4a)$$

where

$$x = \omega_0 t.$$

The latter form (15.4a) is the one which is convenient to use in most applications. Likewise

$$\boxed{b_n = (1/\pi)\int_0^{2\pi} V(t)\cos(nx)\,\mathrm{d}x} \qquad (15.4b)$$

except that the DC term is given by the mean value of $V(t)$:

$$\boxed{b_0 = (1/2\pi)\int_0^{2\pi} V(t)\,\mathrm{d}t.} \qquad (15.4c)$$

A simple application of (15.4a) is to derive the Fourier coefficients in equation (15.2).

Can you understand base-line shift of Chapter 3 in terms of b_0?

In words, what equations (15.4) are saying is that:

$a_n = 2 \times$ *mean value of* $V(t)\sin(n\omega_0 t)$ *over a cycle and*

$b_n = 2 \times$ *mean value of* $V(t)\cos(n\omega_0 t)$ *over a cycle (for* $n \geq 1$*)*

$b_0 =$ *mean value of* $V(t)$ *over a cycle.*

The origin of the factor 2 difference between b_0 and other b_n is the same as the factor 2 difference between DC power $I_0 V_0$ and AC power $\frac{1}{2} I_0 V_0$.

Why are sub-harmonics not present in the Fourier series?

Fourier spectrum

When these coefficients are plotted against n, as in figure 15.4, the histogram is called the **spectrum** of the pulse. Electronic **spectrum analysers** are available commercially to plot out the Fourier components of a waveform. **Synthesisers** do the reverse: they reproduce the sound of an instrument by reconstructing the waveform using oscillators which generate the fundamental and its harmonics using appropriate coefficients and phase relationships. The distinctive character of any musical instrument depends on its Fourier spectrum.

A nice demonstration of Fourier decomposition is to apply square waves to the resonant circuit of figure 14.8. If you adjust the frequency of the square wave close to 1/2 or 1/3 of the resonant frequency, the circuit acts as a narrow pass filter. There

Copyright © 2005 IOP Publishing Ltd.

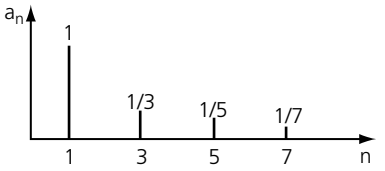


Fig. 15.4. Frequency spectrum of the square waves.

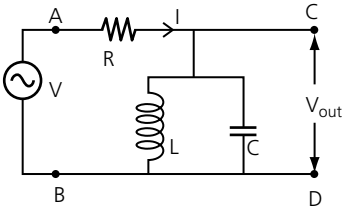


Fig. 14.8. Parallel LC circuit.

are maxima at the output of the filter at these frequencies. The outputs are sine waves at the resonant frequency, i.e. harmonics of the square wave transmitted by the filter. Their magnitudes measure the spectrum shown in figure 15.4. As you vary the frequency away from an exact sub-harmonic of the resonant frequency, you will observe very complicated output waveforms. What is happening is that harmonics are undergoing phase shifts because of the phase variation away from resonance.

If $V(t)$ is symmetrical about $t = 0$, i.e. if $V(t) = V(-t)$, the Fourier series contains only cosine terms and all a_n are zero. If $V(t)$ is antisymmetric about $t = 0$, i.e. $V(t) = -V(-t)$, it contains only sine terms and all b_n are zero. If it is symmetrical above and below the horizontal axis, even harmonics are absent, as in equation (15.2). You should satisfy yourself of those results by writing out special cases from equations (15.4).

15.4 The Hetorodyne Principle

Diodes have characteristic curves which are non-linear. It is instructive to follow through the consequences of this non-linear relation between I and V. To keep the algebra simple, suppose

$$I = aV + bV^2.$$

Next, suppose V is the sum of two AC signals of different frequency:

$$V = V_0 \cos \omega_0 t + V_1 \cos \omega_1 t.$$

Copyright © 2005 IOP Publishing Ltd.

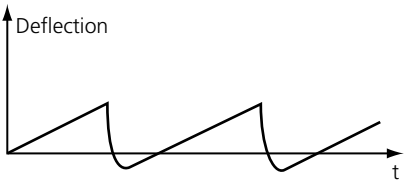


Fig. 15.5. Deflection of a violin string with time.

Then

$$I = a(V_0 \cos \omega_0 t + V_1 \cos \omega_1 t) + b(V_0^2 \cos^2 \omega_0 t + V_1^2 \cos^2 \omega_1 t + 2V_0 V_1 \cos \omega_0 t \cos \omega_1 t).$$

The second term may be rewritten

$$I = \tfrac{1}{2} b(V_0^2 + V_0^2 \cos 2\omega_0 t + V_1^2 + V_1^2 \cos 2\omega_1 t + 2V_0 V_1 \{\cos(\omega_0 - \omega_1)t + \cos(\omega_0 + \omega_1)t\}).$$

In this second term there are outputs at the beat frequencies:

(a) $\omega_0 - \omega_1$
(b) $\omega_0 + \omega_1$
(c) $\omega_0 - \omega_0 = 0$
(d) $\omega_0 + \omega_0 = 2\omega_0$
(e) $\omega_1 - \omega_1 = 0$
(f) $\omega_1 + \omega_1 = 2\omega_1$.

A good audio amplifier is linear and gives very little harmonic distortion. A low-fi radio produces music which is still recognisable because it contains the fundamental frequency ω_0, but it sounds bad because the harmonic spectrum is distorted. Distortion of harmonics confuses one instrument with another.

If the characteristic curve contains a term cV^3, similar algebra shows that a term appears in the output at frequency $3\omega_0$, and so on for higher powers. A very non-linear characteristic curve generates many harmonics.

A musical instrument gives a complicated waveform because the output depends non-linearly on the input driving force. When a bow is pulled across the string of a violin, the string moves in a series of jerks (figure 15.5) as it sticks to the bow; different bowing gives a different pattern of jerks hence a difference in spectrum or timbre.

15.5 Broadcasting

Suppose ω_0 is large (radio frequencies 10^7 rad/s) and ω_1 is small (audio frequencies $\sim 2 \times 10^3$ rad/s for middle C). This might be achieved with the circuit of

Copyright © 2005 IOP Publishing Ltd.

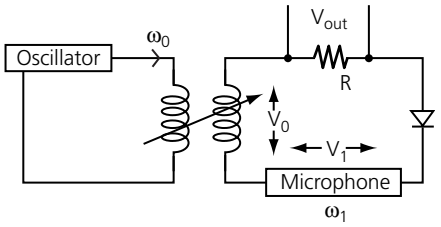


Fig. 15.6. Schematic diagram for broadcasting.

figure 15.6. If the output is fed through a filter which passes frequencies close to ω_0 (as in figure 15.7), the only components which get through the filter are

$$I = aV_0 \cos\omega_0 t + bV_0V_1\{\cos(\omega_0 - \omega_1)t + \cos(\omega_0 + \omega_1)t\}. \tag{15.5}$$

The signals at frequencies $(\omega_0 - \omega_1)$ and $(\omega_0 + \omega_1)$ are called **sidebands**. The expression may be manipulated further:

$$\begin{aligned} I &= aV_0 \cos\omega_0 t + 2bV_0V_1 \cos\omega_0 t \cos\omega_1 t \\ &= aV_0 \cos\omega_0 t\{1 + (2b/a)V_1 \cos\omega_1 t\}. \end{aligned} \tag{15.6}$$

Graphically this has the form shown in figure 15.8. The ratio $2bV_1/a$ is called the **modulation depth** or **modulation index**. The signal contains information about the audio signal of angular frequency ω_1, but the Fourier components of equation (15.5) have been shifted to $(\omega_0 - \omega_1)$ and $(\omega_0 + \omega_1)$. This arrangement is called **amplitude modulation (AM)**. It was the first form used for radio broadcasting at frequencies around $f = 1$ MHz, $\lambda = 300$ m. At frequencies within a factor 5 or so of this, radio waves bounce backwards and forwards between the Earth and the ionosphere; this allows propagation of radio waves over large distances round the Earth. Much shorter or longer wavelengths penetrate the ionosphere.

Suppose the audio signal covers the range up to $\omega_1 = 10^5$ rad/s. The filter of figure 15.7 needs to have a bandwidth twice this, i.e. from say 5.9×10^6 to 6.1×10^6 rad/s. Then one station uses 3% of the available frequency range. Because there are vastly more than 30 stations, a medium-wave receiver picks up beats between the carrier frequencies of different stations, leading to a steady background whine at frequency $\omega_0 - \omega_0'$; this is all too familiar.

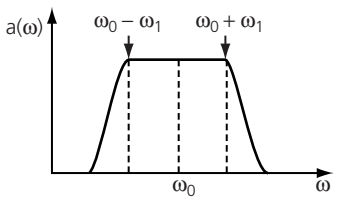


Fig. 15.7. The spectrum passed by a narrow-band filter.

Copyright © 2005 IOP Publishing Ltd.

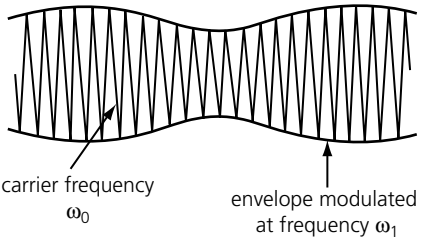


Fig. 15.8. Amplitude Modulation.

Equation (15.5) shows that both sidebands carry the same information. It is possible to suppress one of them and halve the required bandwidth. This is called **Single-Sideband Amplitude Modulation**.

15.6 Frequency Modulation (FM)

An alternative to modulating the amplitude is to keep the amplitude fixed but vary the frequency. Suppose

$$\omega = \omega_0 + B(t) \tag{15.7}$$
$$= \omega_0 + A_f \cos \omega_1 t.$$

The phase of the resulting signal is

$$\phi = \int \omega \, \mathrm{d}t = \omega_0 t + \int B(t') \, \mathrm{d}t'$$
$$= \omega_0 t + \int A_f \cos \omega_1 t \, \mathrm{d}t = \omega_0 t + (A_f/\omega_1) \sin \omega_1 t.$$

If A_f is small,

$$I \simeq \cos \omega_0 t - (A_f/\omega_1) \sin \omega_0 t \sin \omega_1 t \tag{15.8}$$
$$= \cos \omega_0 t + (\tfrac{1}{2} A_f/\omega_1) \left[\cos(\omega_0 + \omega_1)t - \cos(\omega_0 - \omega_1)t\right].$$

The Fourier spectrum is like that in amplitude modulation, except that one sideband is reversed in sign with respect to the carrier. The waveform and its spectrum are shown in figure 15.9.

So far, modulation by a pure cosine term has been considered. More generally the modulating amplitude in equation 15.7 has a complicated time dependence $B(t)$; in the small angle approximation used so far,

$$I = \cos \omega_0 t - A(t) \sin \omega_0 t$$

Copyright © 2005 IOP Publishing Ltd.

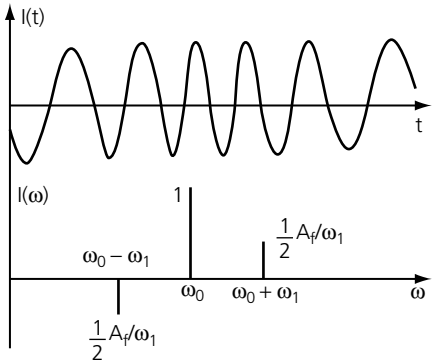


Fig. 15.9. (a) an FM signal, (b) its Fourier spectrum.

where

$$A(t) = \int_0^t B(t')\,\mathrm{d}t'.$$

The signal $A(t)$ can be Fourier analysed, and the second term in equation (15.8) is replaced by its Fourier spectrum. However, it is not essential to restrict the algebra to the approximation that $A(t)$ is small compared to 1. Physically, the wave shown in figure 15.9(a) is perfectly well defined for large $A(t)$, but evaluation of the corresponding expression to equation (15.8), hence the bandwidth, is more elaborate.

If you cut off the top and bottom of the amplitude in figure 15.9(a) at 50% *of the peaks and feed the resulting waveform through a filter centred at* ω_0*, what are the consequences?*

15.7 Frequency Multiplexing

By international agreement, the bandwidth for telephone conversations extends from 300 to 3400 Hz. Most of the spectrum produced by the voice is within this range, and although the removal of high frequencies reduces the quality of the sound, the result is perfectly adequate for conversation.

The bandwidth being used is much less than is available in modern communications. Many separate telephone conversations can be stacked at intervals of 4 kHz in the bandwidth of the link, as indicated in figure 15.10. Individual conversations are modulated on to carrier frequencies spaced by 4 kHz, using a single sideband and suppressing the carrier itself. At the other end of the line, they are recovered by demodulation, i.e. beating against signals at the individual carrier frequencies.

This is called **frequency multiplexing**. The gap of 900 Hz between each conversation and the next is needed for two reasons. Firstly, the filter which isolates one conversation from its neighbours is not perfectly sharp but has rounded

Copyright © 2005 IOP Publishing Ltd.

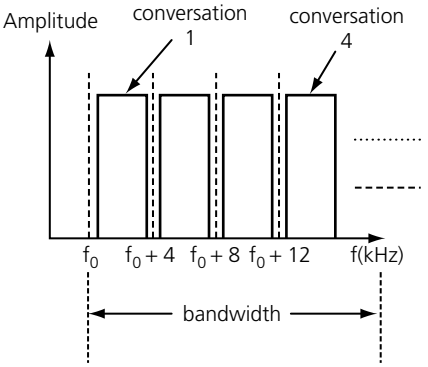


Fig. 15.10. Frequency multiplexing.

edges. Secondly, control information (e.g. on synchronisation and routing) can be transmitted within the gap.

15.8 Time Division Multiplexing

An alternative to sharing the line in frequency intervals is to share it in time. Suppose the signal from one conversation is sampled at regular intervals, figure 15.11(a). A slice of the signal could be despatched down the link and could be reconstituted at the other end by a suitable smoothing scheme. If the sampling is done at a frequency well above 4 kHz and if the bandwidth of the smoothing circuit is limited to 4 kHz, the interpolating curve is limited to 4 kHz and there will be no distortion below this frequency. There is a theorem, due to

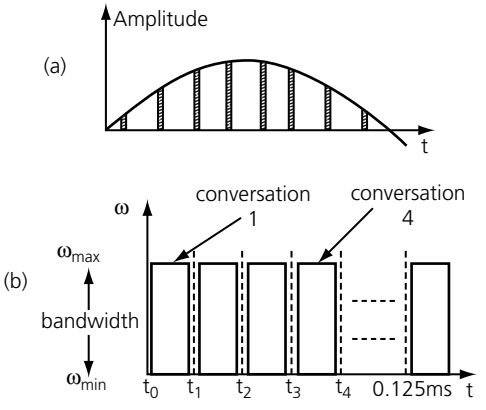


Fig. 15.11. (a) Sampling a waveform, (b) time division multiplexing.

Copyright © 2005 IOP Publishing Ltd.

Shannon, that the signal must be sampled at least twice per cycle, so in practice the sampling is done at 8 kHz.

Slices of individual conversations are transmitted sequentially, figure 15.11(b), using the full bandwidth ω_{max} of the link. Each conversation uses a time interval Δt slightly less than given by $\omega_{max}\Delta t = 1$, so that they do not spread into one another. The input to the link dials through a number of conversations, returning to the first at the sampling frequency of 8 kHz, i.e. after 0.125 ms. This is called **time multiplexing**.

Pulse code modulation

In fact, current practice is not to send the slices themselves but to digitise them. The possible range of amplitudes is broken up into 256 intervals. Thus the amplitude is converted to an 8 bit binary number and it is this number which is transmitted. This is called **Pulse Code Modulation**. Necessarily, there is some truncation in this digitisation, and the result can be in error by up to half an interval. The range of signals from smallest to largest is known as the **dynamic range**.

Noise considerations

There are 8 bits per digitisation and sampling is at 8 kHz, so each conversation requires that bits are transmitted at 64 kHz. Isn't this uneconomical compared with the bandwidth of 4 kHz used in frequency multiplexing?

Frequency multiplexing produces a sideband whose spectrum can vary continuously in amplitude, whereas digital transmission limits the possibilities to 0 or 1. The penalty of the former scheme is that any noise lying at the frequency of the signal will add coherently to it, changing its amplitude or phase from A to $A + \Delta A$. The signal is corrupted. If the attenuation of the signal along the line is frequency dependent, this will also corrupt it. The digital signal, however, is not corrupted unless the noise is so large as to convert 0 to 1 or vice versa. The digital signal will distort progressively along the line from an initial square wave to a rounded shape, as in figure 15.12, because of the bandwidth of the channel. But it can be regenerated at a repeater station before the distortion causes confusion between 0 and 1.

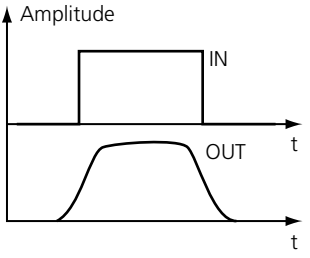


Fig. 15.12. Distortion during propagation.

Copyright © 2005 IOP Publishing Ltd.

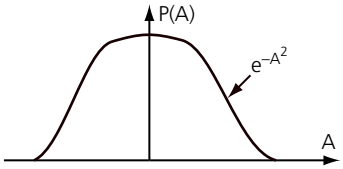


Fig. 15.13. Noise distribution v. amplitude A.

Let us examine this in a little more detail. Suppose the frequency-multiplexed signal is digitised at the downstream end of the link. Suppose it is digitised into n equally spaced intervals; the ratio of noise to interval size depends on $n\Delta A$, where ΔA is the amplitude of the noise. In contrast, for pulse code modulation the noise is ΔA in each of n channels. The total noise power is the same in both cases but is differently distributed. The amplitude of the noise distribution follows a Gaussian curve, figure 15.13, falling as $\exp(-A^2)$ for large amplitude A. Because of the factor n multiplying ΔA, the frequency multiplexed signal is more likely to be corrupted. For TV or telephone signals this is not very important, but for transmission of data from one computer to another it is crucial.

Compact discs digitise the amplitude as 16 bits at a sampling rate of 41.1 kHz. Like satellite links, they include the feature of checking the digitised signal by forming regular sum checks on the recorded signal. By encoding some redundant information, the mistake can usually be corrected.

15.9 Fourier Series using Complex Exponentials

Another important way of writing the Fourier series is using complex exponentials. This is actually simpler than using sines and cosines. Remember that

$$\cos n\omega_0 t = \tfrac{1}{2}(\mathrm{e}^{\mathrm{j}n\omega_0 t} + \mathrm{e}^{-\mathrm{j}n\omega_0 t})$$

$$\sin n\omega_0 t = (1/2\mathrm{j})(\mathrm{e}^{\mathrm{j}n\omega_0 t} - \mathrm{e}^{-\mathrm{j}n\omega_0 t}).$$

It follows that $V(t)$ may be expressed as a series of complex exponentials, but now the series runs from $n = -\infty$ to $+\infty$:

$$V(t) = \sum_{n=-\infty}^{\infty} c_n \mathrm{e}^{\mathrm{j}n\omega_0 t}. \tag{15.9}$$

The Fourier coefficients are easily obtained by multiplying both sides by $\exp(-\mathrm{j}n\omega_0 t)$ and integrating over a cycle:

$$c_n = \frac{\omega_0}{2\pi} \int_{-\pi/\omega_0}^{\pi/\omega_0} V(t)\mathrm{e}^{-\mathrm{j}n\omega_0 t}\,\mathrm{d}t$$

Copyright © 2005 IOP Publishing Ltd.

or

$$\boxed{c_n = \frac{1}{2\pi}\int_{t=-T/2}^{T/2} V(t)\mathrm{e}^{-\mathrm{j}n\omega_0 t}\,\mathrm{d}(\omega_0 t).} \tag{15.10}$$

This replaces equations 15.4 (using sines and cosines) and is simpler. In words, c_n is the mean value of $V(t)\exp -\mathrm{j}n\omega_0 t$ over a cycle. In general, c_n are complex and convey information about both magnitude and phase; two spectra like figure 15.4 are required, one displaying the magnitude and the other the phase of every harmonic.

What is the relation between a_n and b_n of equations (15.4) and c_n?

15.10 Fourier Transforms

The waveform in figure 15.1 was periodic, i.e. it repeated itself after time T. We will now extend Fourier analysis to a single waveform which does *not* repeat. This leads to ideas which are fundamental throughout electrical engineering, and most of physics. Heisenberg's famous Uncertainty Principle is one example.

The requisite formulae can be derived from those we already have by letting $T \to \infty$, in which case the fundamental frequency $\omega_0 \to 0$. The spectrum of frequencies becomes continuous instead of discrete. Suppose we set $\omega_0 = 2\pi/T = \mathrm{d}\omega$, an infinitesimally small quantity, and $n\omega_0 = \omega$; then from the first form of equation (15.10)

$$c_n \to \frac{\mathrm{d}\omega}{2\pi}\int_{-\infty}^{\infty} V(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t = c(\omega)\mathrm{d}\omega \tag{15.11}$$

where

$$c(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} V(t)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t \tag{15.12}$$

and from (15.9)

$$V(t) = \int_{-\infty}^{\infty} c(\omega)\mathrm{e}^{\mathrm{j}\omega t}\,\mathrm{d}\omega. \tag{15.13}$$

The distribution $c(\omega)$ is the spectral representation of $V(t)$ and is mathematically completely equivalent to it. It is called the **Fourier transform** of $V(t)$. It is

Copyright © 2005 IOP Publishing Ltd.

convenient to divide $V(t)$ by $\sqrt{2\pi}$ and rewrite (15.12) and (15.13) in the symmetrical forms:

$$V(\omega) = \sqrt{\frac{1}{2\pi}} \int_{-\infty}^{\infty} V(t) \mathrm{e}^{-\mathrm{j}\omega t} \, \mathrm{d}t. \tag{15.14}$$

$$V(t) = \sqrt{\frac{1}{2\pi}} \int_{-\infty}^{\infty} V(\omega) \mathrm{e}^{\mathrm{j}\omega t} \, \mathrm{d}\omega. \tag{15.15}$$

It is obvious that the same information is embodied in $V(\omega)$ as in $V(t)$, though in a different form; $V(t)$ gives the dependence on t and is said to be **in the time domain**, while $V(\omega)$ gives the dependence on ω and is said to be **in the frequency domain**. Sometimes it is convenient to work with $V(t)$ but in other situations $V(\omega)$ is more convenient, for example when discussing frequency filters. Suppose you want to calculate what happens to a complicated pulse shape like $V(t)$ of figure 15.1 when it goes through a low-pass filter. The steps are: (i) find the Fourier transform of $V_{in}(t)$ using equation (15.14), (ii) multiply the result by the frequency dependence of the filter (in magnitude *and* phase, using a complex transfer function), then (iii) take the Fourier transform of the resulting $V_{out}(\omega)$ to find the output $V_{out}(t)$. By hand this is hopelessly laborious, but using computer packages for Fourier transforms it is easy.

15.11 Response to an impulse

What are Fourier transforms physically? Consider the case when $V(t)$ is an impulse. Imagine, for example, you want to describe a hockey stick hitting a ball. Stick and ball are in contact only very briefly, and we can make the mathematical idealisation that this time interval goes to zero. This is an impulse. It is written $\delta(t)$ (figure 15.14), lasting for an infinitesimally short time Δt, with a height $1/\Delta t$. Integrated over time it gives 1:

$$\int_{-\infty}^{\infty} \delta(t) \, \mathrm{d}t = 1.$$

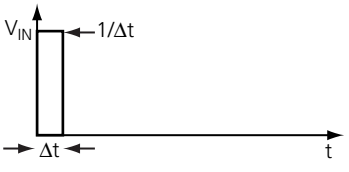


Fig. 15.14. An impulse.

Copyright © 2005 IOP Publishing Ltd.

Its Fourier transform is

$$\begin{aligned}\sqrt{\frac{1}{2\pi}}\int_{-\infty}^{\infty}\delta(t)\mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}t &= \sqrt{\frac{1}{2\pi}}\frac{1}{\Delta t}\int_{0}^{\Delta t}\mathrm{e}^{\mathrm{j}\omega t}\mathrm{d}t \\ &= \sqrt{\frac{1}{2\pi}}\frac{1}{\mathrm{j}\omega\Delta t}(\mathrm{e}^{\mathrm{j}\omega\Delta t}-1) \\ &\to \sqrt{\frac{1}{2\pi}}\frac{\mathrm{j}\omega\Delta t}{\mathrm{j}\omega\Delta t} \qquad \text{as } \Delta t\to 0 \\ &= \sqrt{\frac{1}{2\pi}}.\end{aligned}$$

So the Fourier transform of an impulse is constant at all frequencies: an impulse excites all frequencies equally and with the same phase. We shall find that this holds the clue to relations between frequency response and the time response to a square pulse.

15.12 Fourier Analysis of a Damped Oscillator

Consider the response of a damped oscillator to an impulse at $t = 0$:

$$\begin{aligned}I(t) &= I_0\mathrm{e}^{-t/\tau}\sin\omega_1 t \qquad \text{for } t>0 \qquad (15.16)\\ &= (I_0/2\mathrm{j})\mathrm{e}^{-t/\tau}\{\mathrm{e}^{\mathrm{j}\omega_1 t}-\mathrm{e}^{-\mathrm{j}\omega_1 t}\} \\ &= (I_0/2\mathrm{j})(\mathrm{e}^{s_1 t}-\mathrm{e}^{s_2 t})\end{aligned}$$

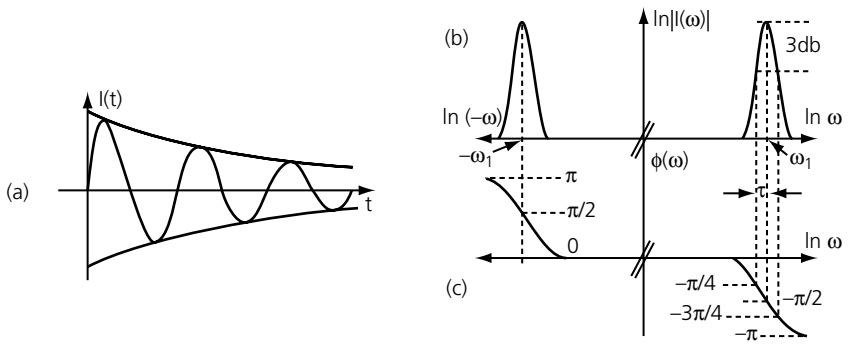


Fig. 15.15. (a) A damped oscillator and (b) and (c) its Fourier transform.

Copyright © 2005 IOP Publishing Ltd.

where

$$s_1 = (-1/\tau) + \mathrm{j}\omega_1$$

$$s_2 = (-1/\tau) - \mathrm{j}\omega_1.$$

Then

$$\begin{aligned} I(\omega) &= \sqrt{\frac{1}{2\pi}}\frac{I_0}{2\mathrm{j}}\int_0^\infty (\mathrm{e}^{s_1 t} - \mathrm{e}^{s_2 t})\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t \\ &= \frac{-I_0}{2\mathrm{j}}\sqrt{\frac{1}{2\pi}}\left[\frac{1}{\mathrm{j}(\omega_1 - \omega + \mathrm{j}/\tau)} - \frac{1}{\mathrm{j}(-\omega_1 - \omega + \mathrm{j}/\tau)}\right] \\ &= \frac{I_0}{\sqrt{8\pi}}\left[\frac{1}{\omega_1 - \omega + \mathrm{j}/\tau} + \frac{1}{\omega_1 + \omega - \mathrm{j}/\tau}\right] \qquad (15.17) \\ &= \frac{I_0\omega_1}{\sqrt{2\pi}}\frac{1}{\omega - \omega_1 - \mathrm{j}/\tau}\cdot\frac{1}{\omega + \omega_1 - \mathrm{j}/\tau}. \end{aligned}$$

Apart from numerical factors, this is just the AC response of the oscillator, equation (14.33), with poles at $\omega = \omega_1 + \mathrm{j}/\tau$ and $\omega = -\omega_1 + \mathrm{j}/\tau$. The waveform $I(t)$ and its Fourier transform are shown in figure 15.15. There are peaks in the magnitude of the spectrum around ω_1 and $-\omega_1$ falling by 3 db when ω moves off the peak by $1/\tau$. Figure 15.15(c) shows the associated phase variation.

The conclusion is that the frequency response is just the Fourier transform of $I(t)$. Physically, an impulse excites all frequencies equally, so what one sees as a function of frequency is just the Fourier transform of the time response to an impulse. The impedance Z depends only on ω and not on time, so if

$$V(t) = Z(\omega)I(t)$$

then

$$\boxed{V(\omega) = Z(\omega)I(\omega).} \qquad (15.18)$$

15.13 The Perfect Filter

A desirable objective is to make an ideal filter, figure 15.16(a), with a flat response from zero frequency up to ω_0 and a sharp cut-off at this frequency. The spectrum transmitted by this circuit will be $h(\omega)g(\omega)$, where $g(\omega)$ is the Fourier transform of the input signal. Unfortunately it turns out to be theoretically impossible. If

Copyright © 2005 IOP Publishing Ltd.

$h(\omega)$ is to have the shape of figure 15.16(a),

$$h(t) = \sqrt{\frac{1}{2\pi}} \int_0^{\omega_0} \mathrm{e}^{\mathrm{j}\omega t}\, \mathrm{d}\omega = \sqrt{\frac{1}{2\pi}} \frac{1}{\mathrm{j}t} \left(\mathrm{e}^{\mathrm{j}\omega_0 t} - 1\right)$$

$$= \sqrt{\frac{1}{2\pi}} \mathrm{e}^{\frac{1}{2}\mathrm{j}\omega_0 t} \frac{\sin \frac{1}{2}\omega_0 t}{\frac{1}{2}t}. \tag{15.19}$$

This requires $h(t)$ to be non-zero for $t < 0$, i.e. the circuit must respond *before* the impulse arrives. This defies common sense, or the so-called **causality principle** that effect follows cause: $h(t) = 0$ for $t < 0$. So it is a physical impossibility to make the perfect filter of figure 15.16(a).

The shape of figure 15.16(a) was square in the frequency domain. Now let's consider a square distribution in the time domain. Suppose a cosine wave lasting for a finite length of time t_0, as in figure 15.16(b):

$$g(t) = \cos \omega_0 t \qquad \text{for} \qquad t = 0 \text{ to } t_0.$$

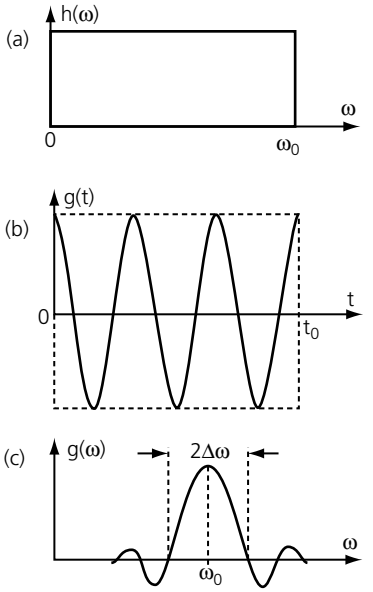


Fig. 15.16. (a) The ideal low-pass filter, (b) a cosine wave of finite duration and (c) its spectrum near $\omega = \omega_0$.

Copyright © 2005 IOP Publishing Ltd.

Then

$$g(\omega) = \tfrac{1}{2}\sqrt{1/2\pi}\int_0^{t_0}\left(\mathrm{e}^{\mathrm{j}\omega_0 t} + \mathrm{e}^{-\mathrm{j}\omega_0 t}\right)\mathrm{e}^{-\mathrm{j}\omega t}\,\mathrm{d}t$$

$$= \sqrt{\frac{1}{2\pi}}\left\{\exp[\mathrm{j}(\omega_0-\omega)t_0/2]\frac{\sin\frac{1}{2}(\omega_0-\omega)t_0}{\omega_0-\omega}\right.$$

$$\left. - \exp[-\mathrm{j}(\omega_0+\omega)t_0/2]\frac{\sin\frac{1}{2}(\omega_0+\omega)t_0}{\omega_0+\omega}\right\}. \qquad (15.20)$$

The spectrum given by equation (15.20) contains two identically shaped peaks around $\omega = \omega_0$ and $\omega = -\omega_0$. One peak is shown in figure 15.16(c). What is happening is that the square time distribution of figure 15.16(b) has a frequency spectrum which produces amplitude modulation of the cosine wave. The half-width $\Delta\omega$ of the peak is given by

$$\tfrac{1}{2}\Delta\omega t_0 = \pi \qquad \text{or} \qquad \Delta\omega = 2\pi/t_0. \qquad (15.21)$$

A narrow spectrum requires large t_0. A small t_0 gives a broad frequency spectrum.

The uncertainty principle

When an atom radiates a light wave, it gets interrupted every now and then by collisions, which alter the phase of the wave in a random way. If the average length of the wave train is $t_0 = \Delta t$, we can substitute $E = h\nu = h\omega/2\pi$ for the energy of the light. From equation (15.21),

$$\Delta E\ \Delta t = h, \qquad (15.22)$$

where h is Planck's constant. This is Heisenberg's uncertainty principle: the energy spread $2\Delta E$ is dictated by the duration Δt of the signal and cannot be less than the value given by equation (15.22). Sometimes Δt and ΔE are defined in terms of RMS values, in which case small but unimportant extra numerical factors appear in equation (15.22).

The message is that Fourier transforms play a central role not only in analysis of electrical circuits but also in quantum mechanics and in optics.

15.14 Exercises

1. *Distortion of a sine wave.* The top and bottom of a sine wave of angular frequency are flattened off by saturation. What effect is there on the frequency specrum?

2. A square wave has the shape shown in figure 15.2. Show that approximately 90% of its intensity is contained in the first and third harmonics

Copyright © 2005 IOP Publishing Ltd.

and that the second harmonic is zero. What modification is required to the square wave to get the second harmonic?

3. A continuous train of square pulses can be analysed in terms of a constant term, a fundamental and harmonics. How is the constant term related to the height of the pulses? Can you see why even harmonics disappear if the width of the pulse is equal to the width of the space between pulses (mark to space ratio 1:1)? Sketch the magnitude of the harmonics if the mark to space ratio is 1:2. What happens as the spacing between pulses increases to a very large value?

Exercises 4 and 5 give some practice in evaluating Fourier coefficients. The algebra is tricky and sometimes lengthy. Once you have the idea, you may subsequently refer to tables for standard results.

4*. Equation (15.2) is the Fourier series for the square waves of figure 15.2, where the signal alternates between $+V_0$ and $-V_0$; verify this result by evaluating the integrals (15.4a) explicitly for $n = 1, 2, 3, 4$ and 5. If instead the square waves are between 0 and 2 V, as in figure 15.17(a), what change is there in the Fourier series? Find the Fourier series for the square waves of figure 15.17(b) by redefining $t = 0$. (Ans: a DC term V_0 is added; $V_0 + (4V_0/\pi)\{\cos\omega_0 t - (1/3)\cos 3\omega_0 t + (1/5)\cos 5\omega_0 t - \ldots\}$.)

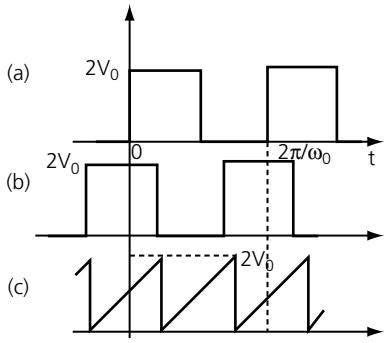


Fig. 15.17.

5*. Show that the waveform in figure 15.17(c) is given for one cycle $x = -\pi$ to $+\pi$ about $t = 0$ by $V = V_0 + V_0 x/\pi$, hence $V(t) = V_0 + (2V_0/\pi)\{\sin\omega_0 t - \frac{1}{2}\sin 2\omega_0 t + \frac{1}{3}\sin 3\omega_0 t - \ldots\}$

6. *Generation of harmonics.* An RF signal of 100 mA at 300 m wavelength is mixed with an audio signal of frequency 1 kHz and magnitude 1 mA by a device having a characteristic $V = AI + BI^2$. The output is passed through a filter with a passband flat between 0.9 and 1.1 MHz and zero elsewhere. What is the output of the filter? What would be the effect of a term CI^3 in the characteristic of the device? (Ans: $V = 0.1A\cos\omega_1 t + 10^{-4}B[\cos(\omega_1 - \omega_2)t + \cos(\omega_1 + \omega_2)t]$

Copyright © 2005 IOP Publishing Ltd.

where $\omega_1 = 2\pi \times 10^6$ and $\omega_2 = 2\pi \times 10^3$ rad/s; $1.5 \times 10^{-7} C\{\cos \omega_1 t + \frac{1}{2}\cos(\omega_1 - 2\omega_2)t + \frac{1}{2}\cos(\omega_1 + 2\omega_2)t\} + 0.5 \times 10^{-3} \cos \omega_1 t$.)

7**. *Amplitude modulation and filtering.* (QMC). For the circuit of figure 15.18, (a) sketch the frequency response, displaying relevant numerical parameters, (b) explain the response to an amplitude modulated signal $V_{in} = (10 + \sin 5 \times 10^4 t)\sin(15 \times 10^4 t)$ mV, (c) explain the response to square waves of angular frequency 5×10^3 rad/s. (Ans: figure 15.19, $(V_{out})_{max} = h_{fe} Z_{max} V_{in}/h_{ie}$ where $Z_{max} = 2 \times 10^5\ \Omega$; (b) $V_{in} = 10 \sin 15 \times 10^4 t + \frac{1}{2}\{\cos 10^5 t - \cos 2 \times 10^5 t\}$: only the signal with $\omega = \omega_0 - 10^5$ has significant amplification; (c) only the harmonic at angular frequency ω_0 has significant gain, so the output is approximately a sine wave at $\omega = \omega_0$.)

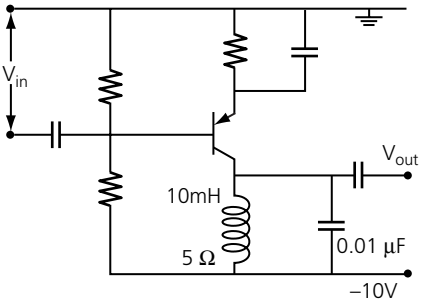


Fig. 15.18.

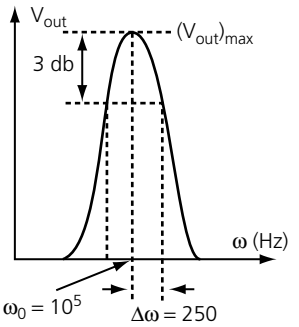


Fig. 15.19.

8. *Demodulation.* The amplitude of a signal $f(t)$ is modulated by multiplying by a carrier wave $A \cos \omega_1 t$. A receiver picks it up with amplitude reduced by a factor B. Show that a replica of the original signal may be recovered by multiplying by $A \cos \omega_1 t$ again, and removing components at angular frequency $2\omega_1 t$ using a filter.

Copyright © 2005 IOP Publishing Ltd.

9. (QMC). *Frequency modulation.* With the aid of time and frequency diagrams, describe frequency modulation. Derive the frequency modulation formula given below and explain each of the terms: $V_{fm}t = A_C \cos(\omega_C t + M \sin \omega_M t)$. An FM transmitter has a bandwidth of 120 kHz. If the carrier frequency is 100 MHz, what is the maximum frequency of the modulating signal if the modulation index M is 5? (Ans: $f_M = 12$ kHz.)

10. *Fourier transform.* A pulse has a shape $f(t)$ as a function of time t and has Fourier transform $F(\omega)$. If the scale of t is increased by a factor α, show that the new Fourier transform is $(1/\alpha)F(\omega/\alpha)$. What is the interpretation of this result?

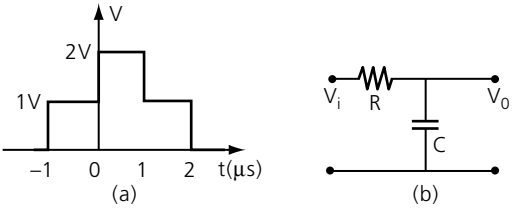


Fig. 15.20.

11. *Fourier transforms.* (QMC). A signal $V = V_0 \exp \mathrm{j}\omega_0 t$ lasts from $t = -\frac{1}{2}\tau$ to $+\frac{1}{2}\tau$ and is zero outside this time interval. Show that its frequency spectrum is $(2V_0/x\sqrt{2\pi}) \sin \frac{1}{2}x\tau$ where $x = \omega - \omega_0$. Show that if it is centred at time t_1 it becomes $V(x) \exp(-\mathrm{j}xt_1)$.

    Hence find the spectrum of the pulse shown in figure 15.20(a). Write down the spectrum of the pulse after it has passed through the filter circuit shown in figure 15.20(b). (Ans: $V'(x) = (8/x\sqrt{2\pi}) \exp(-\frac{1}{2}\mathrm{j}x\tau) \cos^2(\frac{1}{2}x\tau) \sin(\frac{1}{2}x\tau)$ where $\tau = 1\ \mu$s; $V''(x) = V'(x)/(1 + \mathrm{j}\omega CR)$.)

Copyright © 2005 IOP Publishing Ltd.

16

Transformers and 3-Phase Supplies

16.1 Introduction

Electrical power is distributed nationally as AC. At the generator the voltage is typically 20 kV. The National Grid, however, operates at 132 and 270 kV in the UK and at similar values in other countries. This is to reduce power losses (I^2R) in the cables; it keeps cable diameters to reasonable dimensions. The voltage is stepped up from the generator to the Grid by a **transformer**. At the user end of the Grid, a second transformer steps the voltage down again, usually via intermediate 33 and 11 kV stages.

The principle of the transformer is illustrated in figure 16.1. A **primary** coil of n_1 turns is wound on to an iron former, and a **secondary** coil of n_2 turns feeds the output. It will emerge that

$$\frac{\text{secondary voltage}}{\text{primary voltage}} = \frac{\text{n}_2}{\text{n}_1} \quad \text{and} \quad \frac{\text{secondary current}}{\text{primary current}} \simeq \frac{\text{n}_1}{\text{n}_2}.$$

It is easy to demonstrate the first of these relations experimentally using secondary coils with different ratios n_2/n_1. From the second relation, it follows that power is the same in both circuits if the transformer is perfect. That is the essential result. In practice, there is a small loss, typically 1-4%, due to eddy currents in the iron core of the transformer and also because some of the magnetic field generated by the first coil leaves the iron former and does not go through the second coil. If you can imagine a situation where extra coils pick up all the magnetic field and there are no losses, all the power from the first circuit can be transferred to other coils.

Relation between L_1, L_2 and M

In figure 16.1, the two coils have self inductances L_1 and L_2 and mutual inductance M, which will be included explicitly into the circuit diagram of figure 16.2 which follows. Any further impedances in the primary and secondary circuits are lumped into Z_1 and Z_2. These include the losses due to hysteresis in the iron core. Let us first discuss the relation between L_1, L_2 and M, making the assumption that

Copyright © 2005 IOP Publishing Ltd.

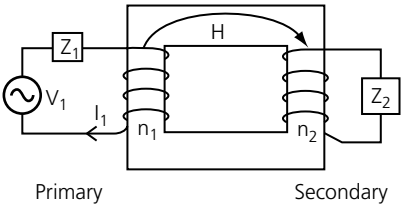


Fig. 16.1. The transformer.

all the magnetic field H circulates through both coils. This is the condition for a **perfect transformer**.

If there are n_1 turns in the primary and n_2 in the secondary

$$H = \alpha_1 n_1 I_1 - \alpha_2 n_2 I_2. \tag{16.1}$$

There is a subtlety of sign here. The way figure 16.1 has been drawn, the coil in the left-hand arm follows a right-handed corkscrew going from the bottom of the figure to the top. The coil in the right-hand side follows a left-handed corkscrew going from bottom to top. Currents I_1 and I_2 therefore contribute to the magnetic field H with opposite signs. The coils could have been wound in either way, but the figure has been deliberately constructed to show that the sign matters.

In equation (16.1) α_1 and α_2 are geometrical constants. We shall not need to know them. The back-emf in circuit 1 is

$$V_1 = \beta_1 n_1 \mathrm{d}H/\mathrm{d}t = \beta_1 n_1(\alpha_1 n_1 \mathrm{d}I_1/\mathrm{d}t - \alpha_2 n_2 \mathrm{d}I_2/\mathrm{d}t)$$

where β_1 is a further geometrical constant. This relation may be written

$$V_1 = L_1 \mathrm{d}I_1/\mathrm{d}t - M_{12}\mathrm{d}I_2/\mathrm{d}t \tag{16.2}$$

with
$$L_1 = \alpha_1 \beta_1 n_1^2 \tag{16.3}$$

$$M_{12} = \alpha_2 \beta_1 n_1 n_2. \tag{16.4}$$

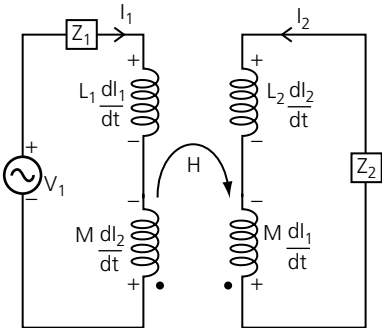


Fig. 16.2. Circuit diagram of Fig. 16.1.

Copyright © 2005 IOP Publishing Ltd.

Likewise, in the second circuit there is a back-emf

$$V_2 = \beta_2 n_2 \mathrm{d}H/\mathrm{d}t = \beta_2 n_2(\alpha_1 n_1 \mathrm{d}I_1/\mathrm{d}t + \alpha_2 n_2 \mathrm{d}I_2/dt)$$

$$= -M_{21}\mathrm{d}I_1/\mathrm{d}t + L_2 \mathrm{d}I_2/\mathrm{d}t \tag{16.5}$$

where $$L_2 = \alpha_2 \beta_2 n_2^2 \tag{16.6}$$

$$M_{21} = \alpha_1 \beta_2 n_1 n_2. \tag{16.7}$$

From (3), (4), (6) and (7),

$$\boxed{M_{12}M_{21} = L_1 L_2} \tag{16.8}$$

for a perfect transformer, and it will be shown below from energy considerations that $M_{12} = M_{21} = M$. Then $M = \pm\sqrt{L_1 L_2}$. The two possible signs for M correspond to the two alternative ways of winding the coils. Frequently this does not matter. The dots in figure 16.1 keep track of the signs of the voltages across L_1, L_2, M_{12} and M_{21}. If the second coil is wound in the opposite sense, M changes sign.

For an **imperfect transformer**, some field produced by circuit 1 does not go through circuit 2 and vice versa. The mutual inductance is due only to that part of the field which goes through both, while the self inductances are due to the whole field through each circuit. For an imperfect transformer

$$\boxed{M^2 = k^2 L_1 L_2 \qquad \text{where} \qquad k^2 < 1.}$$

In this case L_1 is still $\propto n_1^2$, $L_2 \propto n_2^2$ and $M \propto n_1 n_2$.

16.2 Energy Stored in a Transformer

Suppose current I_1 is first established in the primary and then I_2 is brought up from zero, keeping I_1 constant. The work done against the back-emfs is

$$E = \int L_1 \frac{\mathrm{d}I_1}{\mathrm{d}t} I_1 \mathrm{d}t + \int L_2 \frac{\mathrm{d}I_2}{\mathrm{d}t} I_2 \mathrm{d}t - \int M_{12} \frac{\mathrm{d}I_2}{\mathrm{d}t} I_1 \mathrm{d}t$$

$$= \tfrac{1}{2} L_1 I_1^2 + \tfrac{1}{2} L_2 I_2^2 - M_{12} I_1 I_2.$$

The first term corresponds to the power due to current I_1 and the back-emf $L_1 \mathrm{d}I_1/\mathrm{d}t$ in coil 1; likewise for the second term in coil 2. The third term comes from the power in coil 1 due to current I_1 and the back-emf $M_{12}\mathrm{d}I_2/\mathrm{d}t$ due to current I_2.

If the process is carried out in reverse order, first establishing current I_2 in the secondary then bringing up current I_1 from zero in the primary, the indices swop over for M:

$$E = \tfrac{1}{2} L_1 I_1^2 + \tfrac{1}{2} L_2 I_2^2 - M_{21} I_1 I_2.$$

Copyright © 2005 IOP Publishing Ltd.

These two energies must be identical, otherwise a perpetual motion machine could be devised. This requires $M_{12} = M_{21}$. The same result may be obtained from electromagnetic theory by evaluating the geometrical constants α and β. Finally

$$\boxed{E = \tfrac{1}{2}L_1 I_1^2 + \tfrac{1}{2}L_2 I_2^2 - M I_1 I_2.} \tag{16.9}$$

16.3 Circuit Equations and Equivalent Circuits

The circuit analysis now proceeds straightforwardly. For circuits 1 and 2,

$$V_1 = Z_1 I_1 + L_1 \mathrm{d}I_1/\mathrm{d}t - M\mathrm{d}I_2/\mathrm{d}t \tag{16.10}$$

$$M\mathrm{d}I_1/\mathrm{d}t = L_2 \mathrm{d}I_2/\mathrm{d}t + Z_2 I_2. \tag{16.11}$$

These are the fundamental equations giving the response of the two circuits to AC voltages or steps in voltage. They will be manipulated into the form of an equivalent circuit.

Secondary equivalent circuit

Suppose an alternating voltage $V_1 \mathrm{e}^{\mathrm{j}\omega t}$ is applied to the primary. There will be alternating voltages and currents of the same frequency ω in the secondary, and equation (16.11) gives

$$I_2 = \frac{\mathrm{j}\omega M}{Z_2 + \mathrm{j}\omega L_2} I_1. \tag{16.12}$$

Substituting for I_1 in equation (16.10),

$$V_1 = \frac{(Z_1 + \mathrm{j}\omega L_1)(Z_2 + \mathrm{j}\omega L_2)}{\mathrm{j}\omega M} I_2 - \mathrm{j}\omega M I_2$$

$$\text{or} \qquad \frac{V_1}{I_2} = \frac{Z_1 Z_2}{\mathrm{j}\omega M} + \frac{L_1 Z_2}{M} + \frac{Z_1 L_2}{M} + \mathrm{j}\omega \frac{L_1 L_2}{M} - \mathrm{j}\omega M. \tag{16.13}$$

For a perfect transformer, the last two terms cancel and

$$\frac{M}{L_1}\frac{V_1}{I_2} = Z_2 + \frac{Z_1 L_2}{L_1} + \frac{Z_1 Z_2}{\mathrm{j}\omega L_1}.$$

Copyright © 2005 IOP Publishing Ltd.

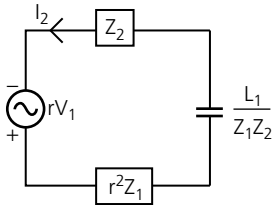


Fig. 16.3. Equivalent circuit of the secondary of a perfect transformer.

Remember that $L_2 \propto n_2^2$, $L_1 \propto n_1^2$ and $M \propto n_1 n_2$, so if the ratio of turns in the secondary and primary is

$$r = n_2/n_1$$

then

$$r\frac{V_1}{I_2} = Z_2 + r^2 Z_1 + \frac{Z_1 Z_2}{\mathrm{j}\omega L_1}.$$

This equation may be interpreted in terms of the equivalent circuit shown in figure 16.3. The driving voltage is rV_1, i.e. V_1 **stepped up** by a factor r. Note carefully that V_1 is the voltage driving the whole primary circuit, not just the voltage across the primary of the transformer. The terms $r^2 Z_1$ and $L_1/Z_1 Z_2$ allow for the voltage drop in the primary across Z_1. If Z_2 is small, equation (16.12) gives

$$\boxed{I_2 \simeq I_1/r;}$$

i.e. the current in the secondary is **stepped down** from that in the primary. In this approximation, $V_1 I_1 = V_2 I_2$ and the transformer simply converts power from one voltage to another.

This point is of practical importance in power systems. However it is also conceptually important. It shows that power can be transmitted through space as well as through a wire; when the transmitted power is absorbed in the secondary, it gives rise to a radiation resistance in the primary circuit. Electromagnetic waves radiated by the primary give rise to radiation resistance.

In 1993, Hulse and Taylor were awarded the Nobel Prize for investigating a corresponding effect in gravitation. Binary pulsars radiate gravitational waves to one another; the absorbtion of these waves slows both pulsars down by resistive losses in the other.

Apart from the capacitor in figure 16.3 (often negligible), the circuit is the Thevenin equivalent of a generator with output impedance $r^2 Z_1$. Thus the transformer changes the matching between the primary voltage source and the load. Small transformers are sometimes used specifically to achieve matching of a high impedance source to a low impedance load (e.g. in driving a loudspeaker, where the standard values of load resistance are 4, 8 and 12 Ω).

What happens to the equations if the secondary coil is wound in the opposite sense to figure 16.1? And what happens to I_2.

Copyright © 2005 IOP Publishing Ltd.

Primary

An equivalent circuit for the primary may be obtained by eliminating I_2 from equation (16.10) using (16.12):

$$V_1 = (Z_1 + \mathrm{j}\omega L_1)I_1 + \frac{\omega^2 M^2 I_1}{Z_2 + \mathrm{j}\omega L_2}$$

$$\text{or} \qquad \frac{V_1}{I_1} = Z_1 + \frac{\mathrm{j}\omega L_1(Z_2 + \mathrm{j}\omega L_2) + \omega^2 M^2}{Z_2 + \mathrm{j}\omega L_2}$$

$$= Z_1 + \frac{\mathrm{j}\omega L_1 Z_2 + \omega^2(M^2 - L_1 L_2)}{Z_2 + \mathrm{j}\omega L_2}. \tag{16.14}$$

Again the term $(M^2 - L_1 L_2)$ disappears for a perfect transformer and

$$\frac{V_1}{I_1} = Z_1 + \frac{\mathrm{j}\omega L_1 Z_2}{Z_2 + \mathrm{j}\omega r^2 L_1} = Z_1 + \frac{\mathrm{j}\omega L_1 Z_2/r^2}{Z_2/r^2 + \mathrm{j}\omega L_1}.$$

This has the equivalent circuit shown in figure 16.4. The load Z_2/r^2 appears across the inductor and is said to be 'reflected' from the secondary. Likewise, in figure 16.3, the load $Z_1 r^2$ is said to be 'reflected' from the primary into the secondary.

Can you understand the appearance of Z_2/r^2 in the primary in terms of energy dissipation in Z_2?

Summary

If a perfect transformer has load Z_2 applied to the secondary where $Z_2 \ll \mathrm{j}\omega L_2$,

$$\boxed{\frac{\text{secondary voltage}}{\text{primary voltage}} = r} \tag{16.15}$$

$$\boxed{\frac{\text{secondary current}}{\text{primary current}} \simeq 1/r} \tag{16.16}$$

and the load seen by the primary $\simeq Z_2/r^2$. The load seen by the secondary is $r^2 Z_1$.

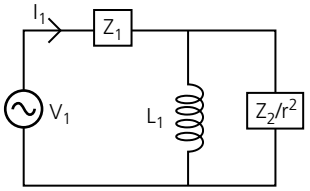


Fig. 16.4. Equivalent circuit of the primary of a perfect transformer.

Copyright © 2005 IOP Publishing Ltd.

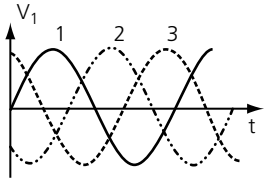


Fig. 16.5. Voltages v. t in a three-phase supply.

Floating AC voltages
There is no direct connection between primary and secondary, so there can be a difference in DC voltage between them. If the primary is earthed, the secondary is said to be floating. Sometimes this is undesirable (e.g. in commercial power supplies) and they are earthed to a common point. It does, however, represent a simple means of eliminating (or introducing) DC biases.

16.4 Three Phase Systems

The AC generators considered so far produce a voltage $V_0 \cos \omega t$ between two terminals. This is what is provided by an ordinary wall-plug between the 'live' and 'neutral'. It suffers the disadvantage that the voltage passes through zero twice per cycle, and so does the power. For some purposes, notably driving an electric motor, it is preferable if the power is more nearly constant. This can be achieved using three supplies differing in phase by 120°, as in figure 16.5. Because of the phase difference of 120°, no two voltages pass through zero simultaneously.

At the generator, this is achieved, figure 16.6, by winding three fixed coils on the **stator** (the stationary part) and providing a rotating magnetic field by means of a **rotor** in the centre. With suitable geometrical design, the magnetic field in each stator coil varies approximately sinusoidally; after some smoothing (as was considered in section 9.5), it induces a sinusoidal voltage in each stator coil. The rotor is supplied with DC via slip-rings. In a power station the rotor is driven by a turbine; in a local generator it is driven from the mains. The design of a

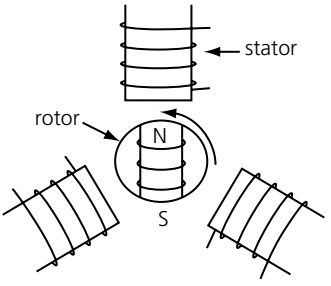


Fig. 16.6. Schematic layout of a three-phase AC generator.

Copyright © 2005 IOP Publishing Ltd.

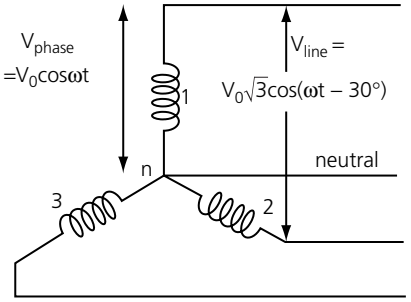


Fig. 16.7. Three-phase supply.

power generator is complicated, because of the large mechanical forces on the conductors, the large energy stored in the rotor and the large voltages and currents.

The three coils are connected together as shown in figure 16.7 with one common terminal, called the **neutral**. This plays two roles. Firstly, it economises on cables by reducing the number of output lines from 6 to 4. Secondly, it avoids unpredictable and unwelcome floating voltages between the three generator coils.

Line voltage

The voltages generated in the three coils vary with time as shown in figure 16.5. On a phasor diagram, their magnitudes and relative phases are as in figure 16.8. The voltage between lines 1 and 2 is $V_{12} = V_1 - V_2$, and if V_2 and V_1 are of equal magnitude V_0, V_{12} lags 30° behind voltage V_1;

$$|V_{12}| = 2V_0 \cos 30° = V_0\sqrt{3}. \tag{16.17}$$

This is known as the **line voltage**. The voltage of one individual phase between live and neutral terminals is known as the **phase voltage**. Household supplies are taken from single phases, live to neutral, i.e. RMS 110 V in the USA, 220 V in Europe and 240 V in the UK; line voltages are a factor $\sqrt{3}$ larger.

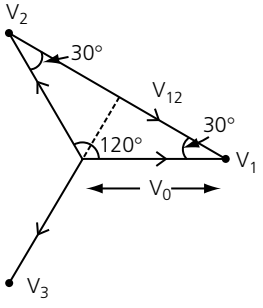


Fig. 16.8. Vector diagram of phase voltages.

Copyright © 2005 IOP Publishing Ltd.

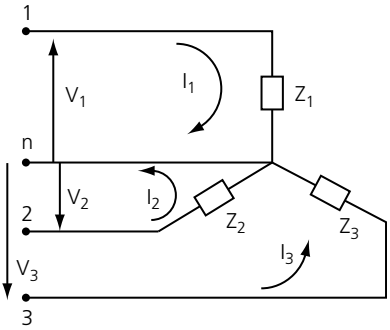


Fig. 16.9. Star connection of loads.

16.5 Balanced Loads

Suppose three loads are connected between live and neutral as shown in figure 16.9. This is known as a **star** configuration. The currents flowing in the three loads are

$$\begin{aligned} I_1 &= V_1/Z_1 \\ I_2 &= V_2/Z_2 \\ I_3 &= V_3/Z_3. \end{aligned} \tag{16.18}$$

If the three loads are the same and equal to Z, the resulting current in the neutral is

$$I_n = I_1 + I_2 + I_3 = \frac{V_1}{Z}(1 + \mathrm{e}^{\mathrm{j}120^0} + \mathrm{e}^{\mathrm{j}240^0}) = 0.$$

This is called a **balanced load**. A three phase motor is normally configured this way. Household supplies are arranged on alternate phases door to door, so as to balance the neutral current approximately.

Suppose your house is supplied by phase 1 and is metered between this phase and neutral. Suppose your neighbour is likewise supplied from phase 2. If you connect a heater between your phase and your neighbour's, are you defrauding (a) the power company, (b) your neighbour?

Power

The total power supplied to the load is

$$P = V_1^2/Z_1 + V_2^2/Z_2 + V_3^2/Z_3. \tag{16.19}$$

In each term, Z is to be interpreted as a complex number:

$$Z = R + \mathrm{j}X = |Z|\mathrm{e}^{\mathrm{j}\phi}$$

and individual terms in equation (16.18) take the form

$$P = V_i^2 \cos\phi_i/|Z_i|. \tag{16.20}$$

Copyright © 2005 IOP Publishing Ltd.

For a resistive load, $Z = R$ and $\cos\phi = 1$. For a purely inductive or capacitative load, $Z = \mathrm{j}X$ and $\cos\phi = 0$, so no power is dissipated; in this case, current and voltage are 90° out of phase. When an electrical motor is running with no load on it, the phase of the rotation adjusts so that voltage and current are 90° apart and no power is drawn. When a load is applied, the motor is retarded so that a phase angle develops between current and voltage and power is absorbed from the mains. The greater the load, the larger the phase retardation.

If the loads are balanced and the neutral voltage is strictly zero, the instantaneous power P is

$$\begin{aligned} P &= (V_0^2/Z)[\cos^2\omega t + \cos^2(\omega t + 120°) + \cos^2(\omega t + 240°)] \\ &= (V_0^2/2Z)\{3 - \cos 2\omega t - \cos(2\omega t + 240°) - \cos(2\omega t + 480°)\} \\ &= 3V_0^2/2Z. \end{aligned} \tag{16.21}$$

The three cosines add up to zero and P is independent of t. This result, that constant power can be supplied, is the essential reason for choosing to deliver power nationally via a three phase system.

Would a 4 or 5 phase supply give a constant power? How about 2 phases? If the mains voltage contains harmonics, what is their effect on the power?

The delta configuration

An alternative way of connecting the loads is shown in figure 16.10. This is called the **delta** configuration. Here the neutral is not connected, and loads are applied directly between live terminals. In this case,

$$I_A = V_{12}/Z_A$$

$$I_B = V_{23}/Z_B$$

$$I_C = V_{31}/Z_C.$$

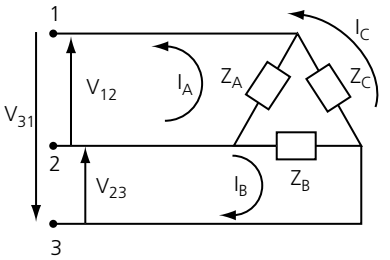


Fig. 16.10. Delta configuration of loads.

Copyright © 2005 IOP Publishing Ltd.

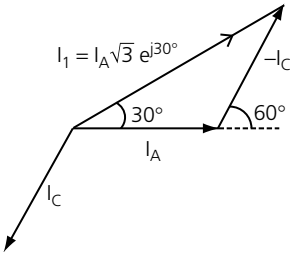


Fig. 16.11. Vector addition of I_1 and I_2.

These are called phase currents. The net current through terminal 1, called the line current, is

$$I_1 = I_A - I_C = (V_{12}/Z_A) - (V_{31}/Z_C).$$

If the loads are balanced and $V_{31} = V_{12}\exp(\mathrm{j}240°)$,

$$I_1 = (V_{12}/Z)(1 - \mathrm{e}^{\mathrm{j}240°});$$

the resultant current is given by the vector diagram of figure 16.11 and is of magnitude $I_A\sqrt{3}$. Thus the line current is $\sqrt{3}$ times the currents through the individual loads. It leads I_A by 30°. If the loads are resistive, it is in phase with the phase voltage, as expected. If the loads are unbalanced, it is necessary to calculate the individual currents I_A, I_B and I_C and add them vectorially.

The power delivered to the loads is

$$P = (V_{12}^2/Z_A) + (V_{23}^2/Z_B) + (V_{31}^2/Z_C)$$

and for a balanced load

$$P = 3|V_{line}|^2/2Z. \tag{16.22}$$

Comparing with equation (16.21), the power delivered to the loads is a factor 3 larger for this configuration than for the star configuration because $V_{line} = V_0\sqrt{3}$, equation (16.17).

16.6 Exercises

1. What are the virtues of 3-phase power distribution? Why is the voltage of the neutral close to earth? What is the relation between line voltage and phase voltage and between line current and phase current? What is the phase relation between line voltage and phase voltage?

   In a transformer, what transmits power between the wires in the primary and the wires in the secondary? For a perfect transformer, why is $L_1L_2 = M^2$?

Copyright © 2005 IOP Publishing Ltd.

2. A star load with three equal arms of 25 Ω is connected across a 3 phase supply with 110 V RMS phase voltage. What is the RMS line voltage? What is the peak current in each arm of the star and the total power dissipation? (Ans: 191 V, 6.22 A, 1.45 kW.)

3. A delta load with three equal arms of $(15 + 20\text{j})$ Ω is connected across a three phase supply with 415 V RMS line voltage. What is the peak phase voltage? What is the RMS line current in magnitude and how does its phase relate to (a) line voltage, (b) phase voltage? (Ans: 339 V, 16.6 A, (a) lags 53.1°, (b) lags 23.1°.)

4. (QMC). A three phase, 415 V system supplies a balanced delta load with impedances of $20\angle -40°$ Ω in each arm and a parallel star load with impedances of $(15 - \text{j}25)$ Ω in each arm. Find the active and reactive power in each load and the magnitude of the total line current. Active power = power dissipated in resistance; reactive power = mean power stored in reactive components. (Ans: 19.8 kW, 16.6 kW, 3.04 kW. 5.07 kW, 34.5 A RMS.)

5. (QMC). Show that the line current is equal to $\sqrt{3}$ times phase current for a three phase delta connected load and draw a phasor diagram for the voltages and currents. A delta connected three phase system has a 10 Ω resistance in series with a 40 mH inductor in each phase of the load. If the load is supplied from a 415 V RMS 50 Hz supply, find the magnitude and phase of the line current and the power supplied to the load. A star connected load, consisting of a 20 Ω resistance in each arm of the star is connected in parallel with the delta load. What is the magnitude and phase of the new line current? (Ans: RMS line current = 25.84 A, 51.5° behind the line voltage; 20.0 kW; 42.0 A, 28.8° behind the line voltage.)

6. A star load with resistances of 25, 30 and 35 Ω is connected between the three phases of a 110 V RMS supply and neutral. What is the RMS magnitude of the current in the neutral? If instead the neutral is left floating, what is its RMS voltage? (Ans: 1.09 A, 10.7 V.)

7. A 415 V three phase supply is connected to the delta network of figure 16.12. What are the RMS currents I_1, I_2 and I_3? (Ans: $I_1 = 15.4$ A, $I_2 = 14.8$ A, $I_3 = 13.0$ A.)

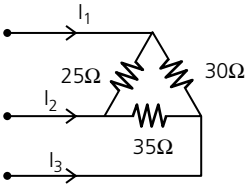


Fig. 16.12.

Copyright © 2005 IOP Publishing Ltd.

8. Figure 16.13 shows the Heavyside bridge. Find the balance conditions and express M in terms of the other components when the bridge is balanced. (Ans: $PR = SQ$ and $R(L_2 + M) = Q(L_3 - M)$.)

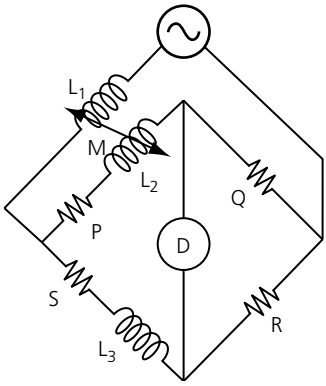


Fig. 16.13.

9. (QMC). Find the magnitude and phase of the current labelled I_1 in figure 16.14. Does it lead or lag the voltage? Sketch the behaviour as a function of frequency. The voltage source V may be taken as $V_0 \cos \omega t$ or $V_0 e^{j\omega t}$ and the transformer may be assumed to be ideal. (Ans: current leads voltage by angle ϕ where $\tan \phi = t = \omega L_1/\{R(\omega^2 C L_2 - 1)\}$; $I_0 = V_0/\{R(1 + t^2)^{1/2}\}$.)

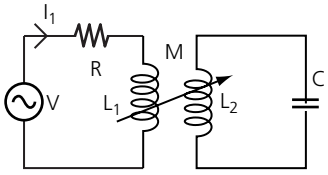


Fig. 16.14.

10. In figure 16.15(a), the transformer turns are in the ratio shown, and the input from a low impedance source is $V_0 \sin \omega t$. Show that the output observed by a high impedance device is $V_0 \sin(\omega t + \phi)$ and find how ϕ varies with R_1. (Ans: Figure 16.15(b).)

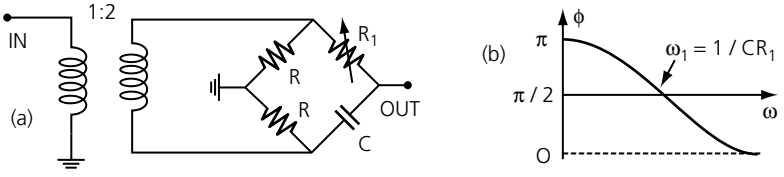


Fig. 16.15.

Copyright © 2005 IOP Publishing Ltd.

11. Use equations 16.10 and 16.11 to show that the two circuits of figures 16.16(a) and (b) are equivalent for suitable values of Z_i, Z_j and Z_k. (Ans: $Z_i = j\omega(L_1 + M)$, $Z_j = -j\omega M$, $Z_k = j\omega(L_2 + M) + Z_2$).

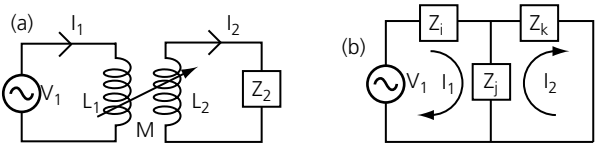


Fig. 16.16.

12. (QMC). Assuming perfect coupling between the primary and secondary windings of the transformer of figure 16.17 and using suitable approximations, calculate the resonant frequency of the primary circuit and the voltage appearing across the 100 mH inductor in the secondary circuit at this resonance. (Ans: There are two possible resonances at $\omega = 10^4\sqrt{10}$ and 10^6 rad/s, but the load shorts the first and only the second survives; 300 mV.)

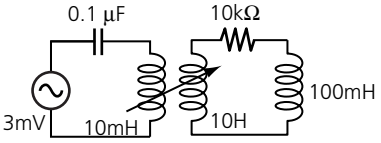
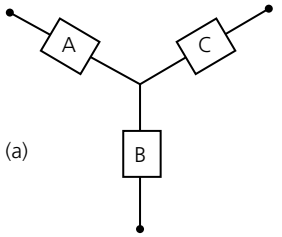


Fig. 16.17.

13. (RHBNC, 1988). Show that the network of figure 16.18(a) is equivalent to that of figure 16.18(b) if $Y_1 = Y_B Y_C/(Y_A + Y_B + Y_C)$, where Y are admittances. Use this for the components in the lower half of the figure to show that the Tuttle bridge of figure 16.19 is balanced if $2\omega^2 LC = 1 - \omega^2 C^2 R_L R = 4R_L/R$.



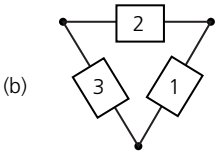


Fig. 16.18.

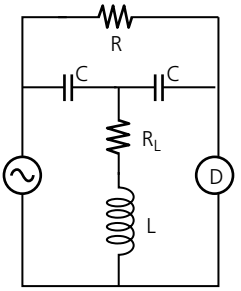


Fig. 16.19.

14*. For an imperfect transformer, $M^2 = k^2 L_1 L_2$. Show from equations (16.12) and (16.13) that the primary and secondary circuits may be

Copyright © 2005 IOP Publishing Ltd.

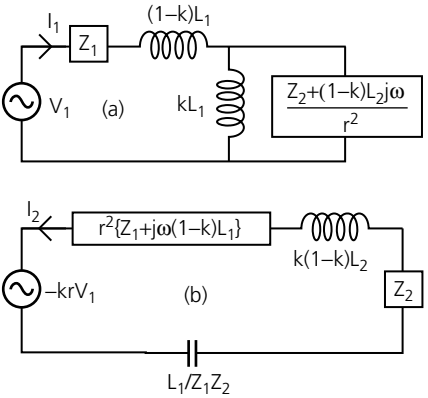


Fig. 16.20.

represented for AC signals of frequency ω by the equivalent circuits shown in figures 16.20(a) and (b).

Copyright © 2005 IOP Publishing Ltd.

Appendix A

Thevenin's Theorem

Any network can be described by loop currents passing from voltage sources to the terminals AB, hence through a load resistor R_L. An example is shown on figure A.1. In order to establish Thevenin's theorem, it is necessary to show that $V_{AB} = V_{EQ} - R_{EQ} I_{AB}$, where V_{EQ} and R_{EQ} do *not* depend on the load resistance R_L.

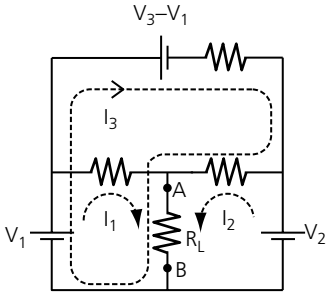


Fig. A1. Loop currents through terminals AB.

Around each current loop, there is a *linear* equation:

$$V_{AB} = V_1 - R_{11} I_1 - R_{12} I_2 - \cdots - R_{1n} I_n$$

$$V_{AB} = V_2 - R_{21} I_1 - R_{22} I_2 - \cdots - R_{2n} I_n$$

$$\vdots$$

$$V_{AB} = V_n - R_{n1} I_1 - R_{n2} I_2 - \cdots - R_{nn} I_n.$$

Because the voltage drops appearing on the right-hand side in the form $R_{ij} I_j$ stop short of the terminals AB, the load resistance R_L does not appear explicitly

Copyright © 2005 IOP Publishing Ltd.

anywhere in the equations. They can be solved for currents (by simple elimination and substitution), with the results

$$I_1 = Y_{11}V_1 + Y_{12}V_2 + \cdots + Y_{1n}V_n - \gamma_1 V_{AB}$$

$$I_2 = Y_{21}V_1 + Y_{22}V_2 + \cdots + Y_{2n}V_n - \gamma_2 V_{AB}$$

$$\vdots$$

$$I_n = Y_{n1}V_1 + Y_{n2}V_2 + \cdots + Y_{nn}V_n - \gamma_n V_{AB}$$

where again none of the coefficients depend on R_L (since R_{ij} did not). Then $I_{AB} = \sum_i I_i$ can be written $B_1V_1 + B_2V_2 + \cdots + B_nV_n - CV_{AB}$. This is of the required form if $C = 1/R_{EQ}$ and $V_{EQ} = (B_1V_1 + B_2V_2 + \cdots + B_nV_n)/C$.

In Chapter 6, it is shown that there is a linear relation between V and I for capacitors and inductors when the notation of complex numbers is used. The linear relation between V and I is the foundation of Thevenin's theorem. Hence the proof carries over to include capacitors and inductors in AC circuits. Fourier's theorem expresses any waveform in terms of AC components, and Thevenin's theorem is therefore generally valid for any network where the relation between V and I is linear. This extends it to small signals in non-linear circuits, following the methods of Chapter 11.

Copyright © 2005 IOP Publishing Ltd.

Appendix B

Exponentials

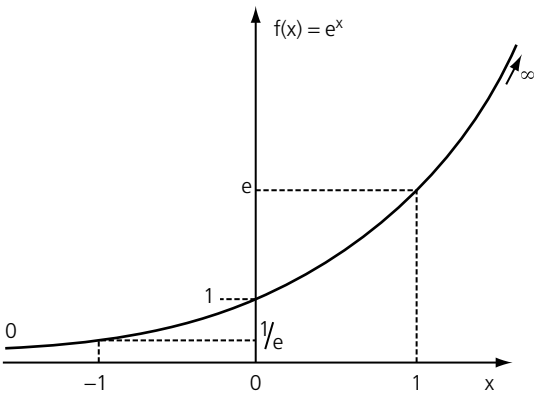


Fig. B1. The exponential function.

The exponential function $f(x) = \mathrm{e}^x$ or $\exp(x)$ is defined by

$$\mathrm{d}f/\mathrm{d}x = f \tag{B.1}$$

$$f(x = 0) = 1. \tag{B.2}$$

It is shown graphically in figure B.1. The slope of the curve is everywhere equal to its value, hence it diverges as $x \to \infty$:

$$\mathrm{e}^{-x} \to \infty \qquad \text{as} \qquad x \to \infty. \tag{B.3}$$

Indeed it diverges rapidly: for $x = 20$, $\mathrm{e}^x = 4.85 \times 10^8$. Even politicians have heard of exponential growth! It will be shown below that $\mathrm{e}^x\mathrm{e}^{-x} = 1$; hence

$$\mathrm{e}^x \to 0 \qquad \text{as} \qquad x \to -\infty. \tag{B.4}$$

The exponential is, however, finite for any finite value of x. It is everywhere positive: if it were to reach the x axis, $\mathrm{d}f/\mathrm{d}x$ would be zero and f would be 0 everywhere.

Copyright © 2005 IOP Publishing Ltd.

With the substitution $x \to \alpha x$ in (B.1)

$$\frac{1}{\alpha}\frac{\mathrm{d}(\mathrm{e}^{\alpha x})}{\mathrm{d}x} = \mathrm{e}^{\alpha x} \qquad \text{or} \qquad \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{e}^{\alpha x}) = \alpha \mathrm{e}^{\alpha x}. \tag{B.5}$$

So, if $\mathrm{d}f'/\mathrm{d}x = \alpha f'$,

$$f' = \mathrm{e}^{\alpha x}. \tag{B.6}$$

Equation (B.5) is important for differentiating exponentials. Conversely,

$$\int \mathrm{e}^{\alpha x}\,\mathrm{d}x = (1/\alpha)\mathrm{e}^{\alpha x} + C \tag{B.7}$$

where C is a constant.

The exponential function may be written as a power series (and is sometimes alternatively defined this way):

$$f(x) = \sum_{n=0}^{\infty} \frac{x^n}{n!} \tag{B.8}$$

$$f(x) = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \frac{x^4}{4!} + \ldots. \tag{B.9}$$

Differentiating this series term by term gives (B.1). Equation (B.1) is also satisfied by

$$f(x) = \text{constant } (1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots).$$

and the condition (B.2) sets the constant to 1. From (B.9) with $x = 1$, e itself is given by 2.718 ...

An important property of the exponential is that

$$\mathrm{e}^x \mathrm{e}^y = \mathrm{e}^{(x+y)}. \tag{B.10}$$

This may be verified by multiplying the power series for e^x and e^y:

$$\begin{aligned}
\mathrm{e}^x \mathrm{e}^y &= (1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \ldots)(1 + y + \frac{y^2}{2!} + \frac{y^3}{3!} + \ldots) \\
&= 1 + x + y + \frac{x^2 + y^2}{2!} + xy + \frac{x^3 + y^3}{3!} + \frac{x^2 y + y x^2}{2!} + \ldots \\
&= 1 + (x + y) + \frac{(x + y)^2}{2!} + \frac{(x + y)^3}{3!} + \ldots \\
&= \mathrm{e}^{(x+y)}.
\end{aligned}$$

Copyright © 2005 IOP Publishing Ltd.

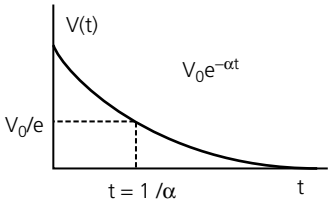


Fig. B2. A signal decaying exponentially with time.

An alternative demonstration is as follows: If $f = e^{\alpha x}$ and $g = e^{\beta x}$,

$$d(fg)/dx = f\,dg/dx + (df/dx)g = \beta fg + \alpha fg \qquad \text{by (B.1)}$$

$$= (\alpha + \beta)fg$$

so fg itself satisfies

$$\frac{d(fg)}{d\{(\alpha+\beta)x\}} = fg$$

hence, from (B.5),

$$fg = e^{(\alpha+\beta)x}.$$

A particularly important exponential in electrical and electronic circuits arises when a signal V (voltage *or* current) decays exponentially with time t:

$$V = V_0 e^{-\alpha t} \equiv V_0 \exp(-\alpha t).$$

The shape of the signal is sketched in Figure B.2. At time $t = 0$, $e^{-\alpha t} = 1$ and $V = V_0$. When $\alpha t = 1$, V has fallen to V_0/e; this value of t is called the **decay time**. As $t \to \infty$, $V \to 0$. The signal obeys the differential equation

$$\frac{dV}{dt} = -\alpha V.$$

Logarithms

The inverse of the exponential is

$$x = \ln_e f. \tag{B.11}$$

From (B.2),

$$\ln_e 1 = 0. \tag{B.12}$$

From (B.1),

$$dx = d(\ln_e f) = df/f \tag{B.13}$$

$$\frac{d}{dx}\{\ln_e f(x)\} = \frac{1}{f}\frac{df}{dx}. \tag{B.13a}$$

Equation (B.13) appears frequently in integrations.

Copyright © 2005 IOP Publishing Ltd.

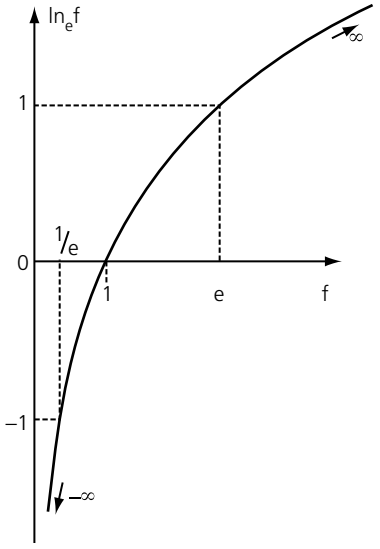


Fig. B3. The logarithmic function.

Figure B.3 shows a sketch of the logarithmic function.

$$\ln_{\mathrm{e}} \mathrm{e} = 1 \tag{B.14}$$

$$\ln_{\mathrm{e}}(1/f) = -x \tag{B.15}$$

$$\ln_{\mathrm{e}} f \to \infty \qquad \text{as} \qquad x \to \infty \tag{B.16}$$

$$\ln_{\mathrm{e}} f \to -\infty \qquad \text{as} \qquad x \to 0. \tag{B.17}$$

Complex Exponentials

If $x = \mathrm{j}\theta$ is substituted in (B.1),

$$\frac{1}{\mathrm{j}}\frac{\mathrm{d}f''}{\mathrm{d}\theta} = f'' \qquad \text{or} \qquad \frac{\mathrm{d}f''}{\mathrm{d}\theta} = \mathrm{j}f''; \tag{B.18}$$

this implies that f'' is complex, and that on the complex plane the slope of f'' is everywhere at 90° to its value. Hence, as θ varies, f'' describes a circle. From the definition (B.2), this is the unit circle. The solution, Figure B.4

$$f'' = \cos\theta + \mathrm{j}\sin\theta \tag{B.19}$$

satisfies (B.5) since

$$\mathrm{d}f''/\mathrm{d}\theta = -\sin\theta + \mathrm{j}\cos\theta = \mathrm{j}f'' = \mathrm{j}\mathrm{e}^{\mathrm{j}\theta}. \tag{B.20}$$

Copyright © 2005 IOP Publishing Ltd.

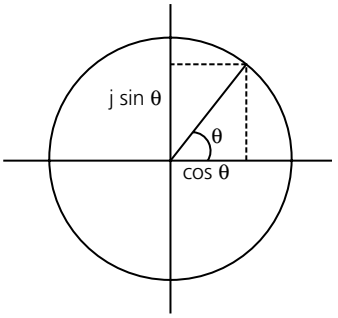


Fig. B4. $e^{j\theta}$.

Complex exponentials satisfy the rule

$$e^{j\theta_1}e^{j\theta_2} = e^{j(\theta_1+\theta_2)} \tag{B.21}$$

as is evident from the product

$$\begin{aligned} e^{j\theta_1}e^{j\theta_2} &= (\cos\theta_1 + j\sin\theta_1)(\cos\theta_2 + j\sin\theta_2) \\ &= (\cos\theta_1\cos\theta_2 - \sin\theta_1\sin\theta_2) + j(\sin\theta_1\cos\theta_2 + \sin\theta_2\cos\theta_1) \\ &= \cos(\theta_1+\theta_2) + j\sin(\theta_1+\theta_2) \\ &= e^{j(\theta_1+\theta_2)}. \end{aligned}$$

Three further useful results are

$$e^{-j\theta} = \cos\theta - j\sin\theta \tag{B.22}$$

$$\cos\theta = \frac{1}{2}(e^{j\theta} + e^{-j\theta}) = 1 - \frac{\theta^2}{2!} + \frac{\theta^4}{4!} - \ldots \tag{B.23}$$

$$\sin\theta = \frac{1}{2j}(e^{j\theta} - e^{-j\theta}) = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \ldots \tag{B.24}$$

Copyright © 2005 IOP Publishing Ltd.